

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**THIS PAGE BLANK (uspto)**



PI Au-Young J;  
 XX  
 DR WPI; 2000-072605/06.  
 DR N-PSDB; AA56719.  
 XX  
 PT proteins, polynucleotides, vectors, host cells and antibodies used to  
 PT diagnose, treat or prevent immune, reproductive, smooth muscle,  
 PT neurological, gastrointestinal, developmental and cell proliferative  
 PT disorders -  
 XX  
 PS Claim 1; Page 128-130; 229pp; English.  
 PS  
 CC AA56698 to AA56776 encode AAY57877 to AAY57955 which represent human  
 CC transmembrane proteins designated HTMPN-1 to HTMPN-79, respectively.  
 CC The transmembrane protein have immunospecific, antiproliferative and  
 CC neuroprotective activities. The human transmembrane proteins,  
 CC polynucleotides encoding them and other compositions and methods from  
 CC the present invention, can be used for the diagnosis, treatment or  
 CC prevention of immune, reproductive, smooth muscle, neurological, or  
 CC gastrointestinal, developmental and cell proliferative disorders. The  
 CC HTMPN's can be used to treat or prevent disorders associated with a  
 CC decreased expression or activity of HTMPN.  
 CC  
 SQ Sequence 688 AA:  
 Query Match 100.0%; Score 3639; DB 21; Length 688;  
 Best Local Similarity 100.0%; Pred. No. 7,4e-234;  
 Matches 688; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 MSASGPGTRRLRLPYMGDLSTQASTQAOQAPANAASTNPPETSNPKPKROT 60  
 DB 1 MSASGPGTRRLRLPYMGDLSTQASTQAOQAPANAASTNPPETSNPKPKROT 60  
 OY 61 NOLQYLLRVLLKTLKHFAMFPQAPVDAVKLNPYYKIKTPMDGTIKRLNNY 120  
 DB 61 NOLQYLLRVLLKTLKHFAMFPQAPVDAVKLNPYYKIKTPMDGTIKRLNNY 120  
 OY 121 NMOECIODENTMTNCTYINKKPGDDIVLMAEALKFLQKINLPLEETIMVQAKGRG 180  
 DB 121 NMOECIODENTMTNCTYINKKPGDDIVLMAEALKFLQKINLPLEETIMVQAKGRG 180  
 OY 181 RGRKETGTAKPGVSYVNTQASTPQOTQOPNPPVQATPHPPFAVNPDLIVQTPVMT 240  
 DB 181 RGRKETGTAKPGVSYVNTQASTPQOTQOPNPPVQATPHPPFAVNPDLIVQTPVMT 240  
 OY 241 VVPQPLQTPRPVPQOPPPAPAPQVQSHPIATPQPVKTKGKVRKADTTPTT 300  
 DB 241 VVPQPLQTPRPVPQOPPPAPAPQVQSHPIATPQPVKTKGKVRKADTTPTT 300  
 OY 301 DPHEPPSLPPEPKTKKIOGRSSRPVKPKKDVDPDSOHPAPEKSKVSEQLKCCSGI 360  
 DB 301 DPHEPPSLPPEPKTKKIOGRSSRPVKPKKDVDPDSOHPAPEKSKVSEQLKCCSGI 360  
 OY 361 LKEMFAKKAAYAMPYKPDVAGLHLYCDIILKHPMDSTIKSLKLEAREYDAOEFGA 420  
 DB 361 LKEMFAKKAAYAMPYKPDVAGLHLYCDIILKHPMDSTIKSLKLEAREYDAOEFGA 420  
 OY 421 DVRLMESNCTIKNPPHEVVAARKIQDYEMFAKMPDEPEEPVAVSSPAVPPTKVV 480  
 DB 421 DVRLMESNCTIKNPPHEVVAARKIQDYEMFAKMPDEPEEPVAVSSPAVPPTKVV 480  
 OY 481 APPSSSDSSSDSSSDSSSTDSSEERAOQLAELQQLAVALHQLALALSPQONKKRKE 540  
 DB 481 APPSSSDSSSDSSSDSSSTDSSEERAOQLAELQQLAVALHQLALALSPQONKKRKE 540  
 OY 541 KDKKK 600  
 DB 541 KDKKK 600  
 OY 601 SEEDKCKPMASYEKKROLSDINKLPGEKLGAVVHIIQSREPSLKNDEIDETIK 660  
 DB 601 SEEDKCKPMASYEKKROLSDINKLPGEKLGAVVHIIQSREPSLKNDEIDETIK 660

OY 661 PSTLREIGALCHILFAEKKETFKRLIM 688  
 DB 661 PSTLREIGALCHILFAEKKETFKRLIM 688  
 RESULT 2  
 AAY07027  
 ID AAY07027 standard; Protein; 754 AA.  
 XX  
 AC AAY07027;  
 XX  
 DT 02-JUL-1999 (first entry)  
 XX  
 DE Breast cancer associated antigen precursor sequence.  
 XX  
 KW Cancer associated antigen; diagnosis; research; treatment; human;  
 KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
 KW prostate cancer.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W09904265-A2.  
 XX  
 PD 28-JAN-1999.  
 XX  
 PF 15-JUL-1998; 98WO-US14679.  
 XX  
 PR 22-JUN-1998; 98US-0102322.  
 PR 17-JUL-1997; 97US-0896164.  
 PR 10-OCT-1997; 97US-0061599.  
 PR 10-OCT-1997; 97US-0061765.  
 PR 10-OCT-1997; 97US-0948705.  
 PR 11-OCT-1997; 97GB-0021697.  
 XX  
 PA (LUDWIG) INST CANCER RES.  
 XX  
 PI Chen Y, Gout I, Gure A, O'Hare M, Ohta Y, Old LJ;  
 PI Pfeundschnuh M, Sahin U, Scanlan MJ, Stockert E;  
 PI Tureci O;  
 XX  
 DR WPI; 1999-132448/11.  
 XX  
 PT New isolated cancer associated nucleic acids and polypeptides -  
 PT isolated using sera from cancer patients; used to develop products  
 PT for the diagnosis, monitoring or treatment of cancers  
 XX  
 PS Disclosure; Page 404-405; 787pp; English.  
 CC  
 CC The invention relates to a method for diagnosing a disorder characterised  
 CC by expression of a human cancer associated antigen precursor coded for by  
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
 CC biological sample isolated from a subject with an agent that specifically  
 CC binds to the NAM, an expression product or a fragment of an expression  
 CC product complexed with an HLA molecule; and (b) determining the  
 CC interaction between the agent and the NAM or the expression product as a  
 CC determination of the disorder. The products and methods can be used in  
 CC the diagnosis, monitoring, research, or treatment of conditions  
 CC characterised by the expression of various cancer associated antigens.  
 CC The invention provides nucleic acid sequences and encoded polypeptides  
 CC which are cancer associated antigen precursors expressed in human breast  
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
 CC lung cancer.  
 CC  
 SQ Sequence 754 AA:  
 Query Match 51.0%; Score 1857; DB 20; Length 754;  
 Best Local Similarity 57.7%; Pred. No. 2.9e-115;  
 Matches 390; Conservative 74; Mismatches 124; Indels 88; Gaps 19;  
 OY 44 NPPPETSNPKPKROTNOLOYLRLVLLKTLKHFAMFPQAPVDAVKLNPYYKIKT 103  
 DB 13 NPPPETSNPKPKRGVNTNLOYLRLKVKMAKMKHGFAMFPQAPVDAVKLNPYYKIKT 72



OY	104	PMDQTKIRRLNNYYNAACIDENFMFNCTYKPGDDIYLMAEALCEFLQKINE	163
Dd	73	PMDQTKIRRLNNTYMAASECDMFTMFNCITYNPDVIDLMAQTLEKITLQKVAS	122
OY	164	LPTETETIMIVQAKC---RG-----RGRKRTGAKDGVSTVPVTTQASTPQOTQPON	214
Dd	133	MQGEQDELVTYIPKNSHKHKAIALAQGSVYSNAHQAVAVSSVSHALTYPPEL-----	166
OY	215	PPPVQATHPPRPAATVPDLIVTPVMTVVPPPOLQTPPPVPPPOQOPAPAPQVQSHPI	274
Dd	187	PTTYLVNIINHP-----SVISSPLIKSLH----SAGPPLLAVTAPA-----	223
OY	275	IATAPQPKTKKGVRKADTPTTIPIIHBP--PSLP--BEKTKTL-GQRESSRPV	328
Dd	224	----QLAKKKGYARKRADTTPTPP-TAIIAPGSPASPQSLEKARALPMRESGRP	277
OY	329	KPKKDVDPDSOHPAPEKSKSYVSQOLCCSGILKEMEFAKKAAYAMPFYKPVDEALGLH	388
Dd	278	KPRKRDLDPDSQQOHOSKKGKLSLDKHCHCNILTELKSKIAATAWPFYKPYDASALGLH	337
OY	389	DYCIDIKKHPMDSTFKSLREAREYRDAQEFQADVRLFMSNCYKNPPDHEVYAMARKLOD	448
Dd	338	DYHDIIKHPMLSTVYKRMENRDYRDAQEFAVDYRLFMSNCYKNPPDHQVAMAARKLOD	397
OY	449	VTFEMFAKMPDEPEEPVAVVSPVAPP-ITYVAPESSSDSSSDSSS-----D	495
Dd	398	VFEFPYAKMPDEPLEPGLPVSTMPPCLAKSSSESSSESSSESSSEEEDEEDDEE	457
OY	496	SDSSTVDSEERAROLAFLQOLRAVHOULALSOPQONPKKKKKKKKKK---EKH	551
Dd	458	ESESSEESSEEAHNLAELQOLRAVHLOLALSOGPISRKRR-KKKKKKKRKAKEH	516
OY	552	KRKEVEENKKS-KAKEPP-PKTKKKNNSNSNVSKKEPA-----PAKSK	594
Dd	517	RGRAGADEDKGPARPRPOPCKKSKKASGSGCSAAIGPSEGFGSGCGTKLPKKAATKTA	576
OY	595	PP--PT-YESEEDCKCMSTEEKROSLDINKLPGEKLGRVHTIOSREPCLKSNPDE	651
Dd	577	PPALPTGTGDSEEBESSRPMZYDERKROLSLDINKLPGERKLGRVHTIOAREPSLDSNPDE	636
OY	652	IEIDPETLKPTLREL	667
Dd	637	IEIDPETLKPTLREL	652
RESULT	3		
ID	AAM81168		
XX	AAM81168 standard; Protein: 947 AA.		
XX			
AC	AAM81168;		
XX			
DT	05-MAR-1999 (first entry)		
XX			
DE	Transcriptional regulatory factor RING3.		
XX			
KW	Human; transcriptional regulatory factor; RING3; TSB; cancer;		
KV	testis specific bromodomain; testicular cell proliferation.		
XX			
OS	Homo sapiens.		
XX			
PN	W09848015-A1.		
XX			
PD	29-OCT-1998.		
XX			
PE	17-APR-1998; 98WO-JP01782.		
XX			
PR	18-APR-1997; 97JP-0116402.		
XX			
PA	(CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.		
XX			
PI	Jones MH;		
XX			
DR	WPI: 1998-583658/49.		

DR	N-PEDB:AAV68343.
XX	Transcriptional regulator gene containing bromodomain sequence - may
PT	be expressed in testis tissue and is useful in treatment of cancer
PT	and other proliferative disorders
XX	
PS	Claim 1: Page 19-24; 42pp; Japanese.
XX	
CC	The present sequence represents the human transcriptional regulatory
CC	factor RING3, which is isolated from testicular cells. RING3 contains
CC	a testis specific bromodomain (TSB) which is expressed specifically
CC	in testis tissue and also expressed in certain tumour lines. The
CC	transgenic cells may be used to express RING3 which is a TSB expression
CC	protein. The TSB expression product can be used in the treatment of
CC	cancer and other proliferative disorders, and in screening of compounds
CC	for ability to bind to it (e.g. for use as drugs by modulation of
CC	transcriptional regulation). DNA capable of hybridising to RING3
CC	polynucleotides may be used for construction of probes and primers.
XX	
SQ	Sequence 947 AA:
Query Match	41.3%; Score 1501.5; DB 19; Length 947;
Best local Similarity	49.8%; Pred. No. 1.7e-91;
Matches 321; Conservative	78; Mismatches 154; Indels 91; Gaps 12.
QY	35 POPANAASTNPPEPETSNNPKRQTNQLOYLRLVYLKTLMKHOFAWPEQOQVDAVKLNL 94
DB	4 PEGQALIVNPPRPPEYINTKKNGRLTNQLOLVKLDLMKHSWPRQRPVDAVKLKL 63
QY	95 PLYYIILTPMDMGITKKRLNNYTNNAOECIODFNMTFNCYIINKREDIVLMAALE 154
DB	64 PLYYIILKPMDLINKRIKLENYKAKASECIEDFNMTFNCYIINKRPDDIVLMAALE 123
QY	155 KFLKINELPFEHEIMVYQAKGGRGKELGTAKRGVSTVPNTTQASTPQTQTPQN 214
DB	124 KLFMKRLSMPDEEOVGV-----KEKIKGT-----QQN 153
QY	215 PPPVQATPPRPVAVPDLIVQTVNTVYVPRQLOTPRPVPRQRPAPARQVOSHPT 274
DB	154 IAVSSAKESPSATEKVFKQOEIPSVFKTSI-----SPLNVQGASV 197
QY	275 IATQAPVKTGKGVRRKADTTPTTIDTPRNEHPSLPREPKTKLQGRSSRPVKR--PK 332
DB	198 NSSQTAQAVTKGVRRKADTTTPAT--SAVKASSEPS-----TTEKSYALPPRIKENMPK 251
QY	333 KVPDSOOHPAPEKSSKVSQOLCCSGILKEKFAKKHAAYAMPFKPVVEALGLHGYCD 392
DB	252 NVLPDSOOQYINVEYKYTEQLRHCSSEILKEKMAKHHSTYAMPFNPDVNALGLHNYTD 311
QY	393 IIKHPDMSIISKLEAREDAOEFGADVRLFMFNSCYKYNPPDIEVVMARKLDVFEK 452
DB	312 VYKNMMDLGTIEKMDNQEKDAYSPADVRLFMFNQCYINPDHVEVYMAAMLDVFEPT 371
QY	453 RFAPKMPDEEP--PVAVASSPAVPPTKYVADPPSSSDSSSDSSSDSDSDSEERAQR 510
DB	372 HFSKPIPIEVEHSPCYIKTD-----TETTGRENTNEASSGNS--DQSEDERVYR 422
QY	511 IALQEOQLKAVHEQLALASQ--PQAKPKKKEKDKKKEK-----HKKEVEENKKS 563
DB	423 IAKLOEOQLKAVHQOQLVLSQVFERLKNKKKEKKEKKEKKNNSNENPRKCCEDMRKE 482
QY	564 KAKEPPPKTKKNNSSNSVSKKEBAPMKSPRPPEYSEEDKCKPMSEEEKROLSDIN 623
DB	483 KSKRNQPKKRRQOFLG-----LKSDEBNAPKPMYDEKROLSLNIN 523
QY	624 KLPGELGRVHHIISREPSLKNSPDEIETFILKSTREL 667
DB	524 KLPGDKLGRVHHIISREPSLNSNPDEIETFILKASTREL 567
RESULT 4	
AAV07114	
ID	AAV07114 standard; Protein: 947 AA.

```

XX AA07114;
AC 02-JUL-1999 (first entry)
DF WO904265 Seq ID No: 685.
DE Cancer associated antigen; diagnosis; research; treatment; human;
KM breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer.
XX Homo sapiens.
XX MO9904265-A2.
XX 28-JAN-1999.
XX 15-JUL-1998; 98MO-US14679.
XX 22-JUN-1998; 98US-0102322.
XX 17-JUL-1997; 97US-0896164.
XX 10-OCT-1997; 97US-0061599.
XX 10-OCT-1997; 97US-0061765.
XX 10-OCT-1997; 97US-0948705.
XX 11-OCT-1997; 97GB-0021697.
XX (LUDW-) LUDWIG INST CANCER RES.
XX Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfrendrich M, Sahlin U, Scanlan MJ, Stockert E;
PI Threlk O;
XX WPI: 1999-132448/11.
XX New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX Disclosure: Page 728-730; 787pp; English.
XX The invention relates to a method for diagnosing a disorder characterised
XX by expression of a human cancer associated antigen precursor coded for by
XX a nucleic acid molecule (NAM). The method comprises: (a) contacting a
XX biological sample isolated from a subject with an agent that specifically
XX binds to the NAM, an expression product or a fragment of an expression
XX product complexed with an HLA molecule; and (b) determining the
XX interaction between the agent and the NAM or the expression product as a
XX determination of the disorder. The products and methods can be used in
XX the diagnosis, monitoring, research, or treatment of conditions
XX characterised by the expression of various cancer associated antigens.
XX The invention provides nucleic acid sequences and encoded polypeptides
XX which are cancer associated antigen precursors expressed in human breast
XX cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
XX lung cancer.
XX Sequence 947 AA:
XX
XX Query Match 41.3%; Score 1501.5; DB 20; Length 947;
XX Best Local Similarity 49.8%; Pred. No. 1.7e-91;
XX Matches 321; Conservative 78; Mismatches 154; Indels 91; Gaps 12;

```

```

QY 215 PPPVQATPHPPFAVTPDILIVQTPVMTVPPQPLQTPPPVPPQPPAPAPQVOSHPI 274
DB 154 IAVSSAKKSSPSATEKVKQOEIPSVFPKTSI-----SPINVVQGASV 197
QY 275 IATPQPVKTKGKVRKADTTPTTIDPIHEPPSLPPEPKTKTKLGORRESSRPV--PK 332
DB 198 NSSQTAQVTVGVRKADTTTPAT-SAVKASSESP-----TTEKSVALLPPIENMPK 251
QY 333 KDVPSQQAPEKSKVSEQLKCCGSLIKEMFARKHAAYAMPFYKPYDVEALGLHDYCD 392
DB 252 NVLPDSQQQYVNVETVKVTEQLRHCSSELLKEMAKKHSYAMPFYNPDVNALGLHNYD 311
QY 393 ITKHPMDKSTISKLEARYRDAQERGADYRLAFNSCYKYNPPDHEVYVAMAKKLDVFEK 452
DB 312 VYKNPMDLGTIEKMDNQYKDAYSPADYRILFMNCKYKYNPPDHEVYVAMAKKLDVFEK 371
QY 453 RFKAMPDEPEE--PYAAVSPAVPPTKVAAPSSSDSSSDSSSDSDSDTDESEERQR 510
DB 372 HFSKIPTEPESMPLCYITD-----TETTGRENTNEASGNS-DSSEDERYKR 422
QY 511 LAELQELKVAHEQLAALSQ-PQONKPKKKKKKKKKKK-----HKRKEVEENKKS 563
DB 423 LAKLQELKVAHQQLQVLSQVFRKLKKKKKKKKKKKKNNNNENNRKKCEQMRKE 482
QY 564 KAKEPPPKTKKNNNSNSNVSKKEPAPKSKPPPYESEEDKCKPMSTEERKQLSLDIN 623
DB 483 KSKRNPQKKRKOQFTG-----LKSEEDENAKPMYDEKROLNLIN 523
QY 624 KLPGKLGKRVHIIOSRPSLKNPNDEIETLKPSTLREL 667
DB 524 KLPGKLGKRVHIIOSRPSLKNPNDEIETLKPSTLREL 567

```

RESULT 5  
ABB58985  
ID ABB58985 standard; Protein; 1937 AA.  
XX ABB58985;  
XX  
XX 26-MAR-2002 (first entry)  
XX  
XX Drosophila melanogaster polypeptide SEQ ID NO 3747.  
XX Drosophila: developmental biology; cell signalling; insecticide;  
XX pharmaceutical.  
XX Drosophila melanogaster.  
XX WO200171042-A2.  
XX  
XX 27-SEP-2001.  
XX  
XX 23-MAR-2001; 2001WO-US09231.  
XX  
XX 23-MAR-2000; 2000US-191637P.  
XX 11-JUL-2000; 2000US-0614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX Venter JC, Adams M, Li PWD, Myers EW;  
XX WPI: 2001-656860/75.  
XX N-PSDB: ABL03088.  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
XX genes from Drosophila and for elucidating cell signalling and cell-cell  
XX interactions -  
XX Disclosure: SEQ ID NO 3747; 21pp + Sequence Listing; English.  
XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and



PR	11-JUL-2000	2000US-0217486
PR	11-JUL-2000	2000US-0217496
PR	14-JUL-2000	2000US-0218290
PR	26-JUL-2000	2000US-0220963
PR	26-JUL-2000	2000US-0220964
PR	14-AUG-2000	2000US-0224518
PR	14-AUG-2000	2000US-0224519
PR	14-AUG-2000	2000US-0225213
PR	14-AUG-2000	2000US-0225214
PR	14-AUG-2000	2000US-0225266
PR	14-AUG-2000	2000US-0225267
PR	14-AUG-2000	2000US-0225267
PR	14-AUG-2000	2000US-0225270
PR	14-AUG-2000	2000US-0225447
PR	14-AUG-2000	2000US-0225457
PR	14-AUG-2000	2000US-0225758
PR	14-AUG-2000	2000US-0225759
PR	18-AUG-2000	2000US-0226579
PR	22-AUG-2000	2000US-0226681
PR	22-AUG-2000	2000US-0226688
PR	23-AUG-2000	2000US-0227182
PR	23-AUG-2000	2000US-0227009
PR	30-ANG-2000	2000US-0229824
PR	01-SEP-2000	2000US-0229287
PR	01-SEP-2000	2000US-0229343
PR	01-SEP-2000	2000US-0229344
PR	01-SEP-2000	2000US-0229345
PR	05-SEP-2000	2000US-0229509
PR	05-SEP-2000	2000US-0229513
PR	06-SEP-2000	2000US-0230437
PR	06-SEP-2000	2000US-0230438
PR	08-SEP-2000	2000US-0231442
PR	08-SEP-2000	2000US-0231443
PR	08-SEP-2000	2000US-0231444
PR	08-SEP-2000	2000US-0231413
PR	08-SEP-2000	2000US-0231414
PR	08-SEP-2000	2000US-0232080
PR	08-SEP-2000	2000US-0232081
PR	12-SEP-2000	2000US-0231968
PR	14-SEP-2000	2000US-0232397
PR	14-SEP-2000	2000US-0232398
PR	14-SEP-2000	2000US-0232399
PR	14-SEP-2000	2000US-0232401
PR	14-SEP-2000	2000US-0232401
PR	14-SEP-2000	2000US-0233063
PR	14-SEP-2000	2000US-0233064
PR	14-SEP-2000	2000US-0234223
PR	21-SEP-2000	2000US-0234223
PR	21-SEP-2000	2000US-0234274
PR	25-SEP-2000	2000US-0234997
PR	25-SEP-2000	2000US-0234998
PR	26-SEP-2000	2000US-0235484
PR	27-SEP-2000	2000US-0235534
PR	27-SEP-2000	2000US-0235636
PR	29-SEP-2000	2000US-0236527
PR	29-SEP-2000	2000US-0236528
PR	29-SEP-2000	2000US-0236568
PR	29-SEP-2000	2000US-0236569
PR	29-SEP-2000	2000US-0236570
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-0237038
PR	02-OCT-2000	2000US-0237039
PR	13-OCT-2000	2000US-0237040
PR	13-OCT-2000	2000US-0239335
PR	13-OCT-2000	2000US-0239337
PR	20-OCT-2000	2000US-0240960
PR	20-OCT-2000	2000US-0241221
PR	20-OCT-2000	2000US-0241785
PR	20-OCT-2000	2000US-0241787
PR	20-OCT-2000	2000US-0241808
PR	20-OCT-2000	2000US-0241809

PR	20-OCT-2000;	2000US-0241825.
PR	01-NOV-2000;	2000US-0244617.
PR	08-NOV-2000;	2000US-0246474.
PR	08-NOV-2000;	2000US-0246475.
PR	08-NOV-2000;	2000US-0246476.
PR	08-NOV-2000;	2000US-0246477.
PR	08-NOV-2000;	2000US-0246478.
PR	08-NOV-2000;	2000US-0246523.
PR	08-NOV-2000;	2000US-0246524.
PR	08-NOV-2000;	2000US-0246525.
PR	08-NOV-2000;	2000US-0246526.
PR	08-NOV-2000;	2000US-0246527.
PR	08-NOV-2000;	2000US-0246528.
PR	08-NOV-2000;	2000US-0246532.
PR	08-NOV-2000;	2000US-0246609.
PR	08-NOV-2000;	2000US-0246610.
PR	08-NOV-2000;	2000US-0246611.
PR	08-NOV-2000;	2000US-0246613.
PR	17-NOV-2000;	2000US-0249207.
PR	17-NOV-2000;	2000US-0249208.
PR	17-NOV-2000;	2000US-0249209.
PR	17-NOV-2000;	2000US-0249210.
PR	17-NOV-2000;	2000US-0249211.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249213.
PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249264.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249297.
PR	17-NOV-2000;	2000US-0249299.
PR	17-NOV-2000;	2000US-0249300.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251988.
PR	05-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251988.
PR	08-DEC-2000;	2000US-0251997.
PR	11-DEC-2000;	2000US-0254097.
PR	05-JAN-2001;	2001US-0259678.
XX	(HUMA-) HUMAN GENOME SCI INC.	
XX		
XX	Rosen CA,	Barash SC, Ruben SM;
XX		
XX	WPI; 2001-488783/53.	
DR	N-PSDB; AAS26193.	
DR		
XX		
PT	New nucleic acid molecules encoding 461 human secreted proteins for	
PT	diagnosing, preventing, treating or ameliorating medical conditions and	
PT	used as food additives or preservatives -	
XX		
XX		
PS	Claim 11; SEQ ID No 1159; 980pp; English.	
XX		
CC	The invention relates to isolated nucleic acid molecules and their	
CC	encoded secreted proteins. The nucleic acids and proteins are used to	
CC	prevent, treat or ameliorate a medical condition in e.g. humans, mice,	
CC	rabbits, goats or horses, cats, dogs, chickens or sheep. They	
CC	are also used in diagnosing a pathological condition or susceptibility	
CC	to a pathological condition. Antibodies to the proteins can also	
CC	be used in alleviating symptoms associated with the disorders and in	
CC	diagnostic immunoassays e.g. radioimmunoassays or enzyme linked	
CC	immunosorbant assays (ELISA). Disorders which are diagnosed or treated	

Query Match	16.3%	Score 593;	DB 22;	Length 235;
Best Local Similarity	54.1%;	Pred. No. 8.9e-32;		
Matches 124;	Conservative 21;	Mismatches 56;	Indels 28;	Gaps 5;
QY 51	SNPNPKRQTNLOLRLVRLKTLKHOFAPFOOPVAVKLTLPDYKIKITPMDMGTI	110		
Db 1	SNPSRGKRTNLOLQMONVVKTLKHOFAPFPYOPVDAIKLTLPDYKIKITPMDMGTI	60		
QY 111	KRLLENNTYWNAAOEIODEFTWFTNCYIYKNGGDIYVMAALEKLFLOKINELPTEETE	170		
Db 61	KRLLENNTYWNAAOEIODEFTWFTNCYIYKNGGDIYVMAALEKLFLOKINELPTEETE	120		
QY 171	IMTVQAKGGRGKRTGTAKPGVSTVPMTQASRPPQOTQOPNPDPVQATPHPPAVTP	230		
Db 121	ITTPCSKG-----XKVGSRLEPRPAOVHSMKRCPLSPQRPFRACPPSPRPRS	170		
QY 231	DLIVQPTWTVTP--QP-LQF-----DPVPBPQGPBPAPAPQVQSH	271		
Db 171	SL-----PPLXPSLQTSRKSQSPQLPRLHLLPSSPWSLNR	209		
RESULT 8				
AAU16619				
.ID	AAU16619	standard;	Protein;	140 AA.
XX AC	AAU16619;			
XX DT	07-NOV-2001	(first entry)		
XX XX	Human novel secreted protein, Seq ID 1572.			
DE				
XX	Human; immunosuppressive; antiarthritic; antirheumatic;			
KM	cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;			
KM	neuroprotective; antibacterial; vlnicide; fungicide; ophthalmological;			
KM	vulnerrary; secreted protein; Rheumatoid arthritis;			
KM	vulnerrary; secreted protein; Rheumatoid arthritis;			
KM	cardiovascular disorder; cardiovascular disorder; cardiac arrest;			
KM	cardiovascular disorder; Alzheimer's disease; angiogenesis;			
KM	neovous system disorder; Alzheimer's disease; infection; ocular disorder;			
KM	corneal infection; wound healing; epithelial cell proliferation;			
XX	skin ageing; food additive; preservative; antiproliferative.			
OS				
XX	Homo sapiens.			
XX				
PN	WO200155322-A2.			
XX				
PD	02-AUG-2001.			
XX				
PF	17-JAN-2001; 2001WO-US01341.			
XX				
PR	31-JAN-2000; 2000US-0179065.			
PR	04-FEB-2000; 2000US-0180628.			
PR	24-FEB-2000; 2000US-0184664.			
PR	02-MAR-2000; 2000US-0186350.			
PR	16-MAR-2000; 2000US-0189874.			
PR	17-MAR-2000; 2000US-0190076.			
PR	18-APR-2000; 2000US-0198123.			
PR	19-MAY-2000; 2000US-0205515.			



CC are also used in diagnosing a pathological condition or susceptibility  
CC to a pathological condition. Antibodies to the proteins can also  
CC be used in alleviating symptoms associated with the disorders and in  
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated  
CC include autoimmune diseases e.g. rheumatoid arthritis,  
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.  
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi  
CC and ocular disorders e.g. corneal infection, and many other  
CC disorders listed in the specification. The polypeptides can also  
CC be used to aid wound healing and epithelial cell proliferation, to  
CC prevent skin aging due to sunburn, to maintain organs before  
CC transplantation, for supporting cell culture of primary tissues, to  
CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
CC as a food additive or preservative to increase or decrease storage  
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
CC minerals, cofactors and other nutritional components. The present  
CC sequence represents a novel secreted protein of the invention.

Query Match 15.5%; Score 564; DB 22; Length 140;  
Best Local Similarity 75.4%; Pred. No. 4,1e-30;  
Matches 101; Conservative 16; Mismatches 17; Indels 0; Gaps 0;

QY 51 SNPKPKRQTNQQLYLRVLTLMKHQFAMPQOPVDVAVLNPYYKIKTPMDGTI 110  
DB 1 SNPKPKRQTNQQLYLRVLTLMKHQFAMPQOPVDVAVLNPYYKIKTPMDGTI 60  
QY 111 KKRLNNYVMAOECIODENFTNCTYNNKPGDIYVMAALRKLFLQKINELPTETE 170  
DB 61 KKRLNNYVMAOECIODENFTNCTYNNKPGDIYVMAALRKLFLQKINELPTETE 120  
QY 171 IMIVQAKGRGRK 184  
DB 121 IITPCSKGAKYGR 134

RESULT 9  
ABB68341.  
ID ABB68341 standard; Protein; 513 AA.

AC ABB68341:  
XX 26-MAR-2002 (first entry)  
DE Drosophila melanogaster polypeptide SEQ ID NO 31815.

KW Drosophila; developmental biology; cell signalling; insecticide;  
pharmaceutical.

OS Drosophila melanogaster.

XX MO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI, 2001-656860/75.

XX DR N-PSDB; ABL12444.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -

PS Disclosure; SEQ ID NO 31815; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB161716-AB130511), expressed DNA  
CC sequences (AB101840-AB161715) and the encoded proteins  
CC (AB157737-AB172072).

XX The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

QY 26 MSTQAOA-QPQANMASTNP-----PPETSNPKPKRQTNQQLYLRVLTLMKHQF 79  
DB 1 MNELOSNQPPRPNRPYLOPVNGVOPVYPPRNRGRNRNILE-LKSYLNCMLNRKF 59  
Matches 151; Conservative 76; Mismatches 229; Indels 178; Gaps 21;

QY 80 AMPFOQPVDAVKLNPYYKIKTPMDGTIKRLNNYVMAOECIODENFTNCTY 139  
DB 60 SYHFRHPDVSLSGDPVHAHVKKHMDLSTIRKRLHNKYVQASBALDFKILINDCLLY 119  
QY 140 NKPGDDIYVMALEKFLQKIN-ELPTEETIMIVQAKGRGRKGTG- 188  
DB 120 NLESGPYVQAGKLMEAFYMRSEIDISTE-----VELKPSKRRKATATESDQASTS 173

QY 189 -AKPGVSTVPWTQASFPQOTPOPNPPOVQA-----TPHPPAVNPDLIVOPV 238  
DB 174 FSAPRASNNYRQWLSSSSMLCPRPMGVSFRNPRNRYVGSILPSPFMDSLV-NPM 232

QY 239 MTVPVPOPLQTPP-----PVBPQOPPPAPAPQVSHPPPIAATPOPVTKKGVRKAD 293  
DB 233 QSMHPMSMNPPIFKNNMETNEADPPPS---EPISYRPLDSLAPLP- 277

QY 294 TTPPTTIDPIHEPSPLP-PEPKTKLGGRRSSRPVPPKRDVDSOHPAP- 345  
DB 278 -----SPMEPTLTPPTP-----APVESPASSP-----PAPNPPIIICY 313

QY 346 KSKVSEQLKCCSGILKEMFAKHAAYAMP-----YKPYDVEALGLHYCDIK 395  
DB 314 KSLDRMIEKSHDHLMSVKKRKYQYTAFOADYMRRSQNP-----DYDHDRE 364

QY 396 HPMDSTIKSLAREYDAQDFGADYRLMFSNCTYKYPDPHEVYVAMARKLQDFEMRFA 455  
DB 365 EKLDMKILQERLSDNFESFDFGVSVKMKMQLRCPEDEGLVKSYSKKTNEIFERL 424

QY 456 KMPDEEPPEVYVAVSSPAVPPTKYVAPPSSSSSSSSSDSDSDSTDSEERARLAELO 515  
DB 425 K-----YRELI 430

QY 516 EQLKAVHEOLAALSOPOONKPKKKKKKKKKKKRKEVEENKKSKAKPPPKTKK 575  
DB 431 ATAKEGRQLVA-----SREODFRSDNLILK-QENANNNKFGWPKPTIITES 479

QY 576 NNSNSNVSKKEPAPMKSKPP-----PTYESE 602

DB 480 LFOVQORVYKMEPDQVOEFSPGTEPLDQYVDSO 513

RESULT 10  
AAG41763  
ID AAG41763 standard; Protein; 568 AA.

XX AAG41763:  
XX 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 52002.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EPI033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 990S-0121825.  
PR 09-MAR-1999; 990S-0123180.  
PR 05-MAR-1999; 990S-0123548.  
PR 23-MAR-1999; 990S-0125788.  
PR 25-MAR-1999; 990S-0126264.  
PR 29-MAR-1999; 990S-0126785.  
PR 01-APR-1999; 990S-0127462.  
PR 06-APR-1999; 990S-0128234.  
PR 08-APR-1999; 990S-0128714.  
PR 16-APR-1999; 990S-0129845.  
PR 19-APR-1999; 990S-0130077.  
PR 21-APR-1999; 990S-0130449.  
PR 23-APR-1999; 990S-0130510.  
PR 28-APR-1999; 990S-0130891.  
PR 30-APR-1999; 990S-0131449.  
PR 30-APR-1999; 990S-0132048.  
PR 04-MAY-1999; 990S-0132407.  
PR 04-MAY-1999; 990S-0132484.  
PR 05-MAY-1999; 990S-0132485.  
PR 06-MAY-1999; 990S-0132486.  
PR 07-MAY-1999; 990S-0132487.  
PR 11-MAY-1999; 990S-0132863.  
PR 14-MAY-1999; 990S-0134218.  
PR 14-MAY-1999; 990S-0134219.  
PR 14-MAY-1999; 990S-0134221.  
PR 14-MAY-1999; 990S-0134370.  
PR 18-MAY-1999; 990S-0134768.  
PR 19-MAY-1999; 990S-0134941.  
PR 20-MAY-1999; 990S-0135124.  
PR 21-MAY-1999; 990S-0135353.  
PR 24-MAY-1999; 990S-0135629.  
PR 25-MAY-1999; 990S-0136021.  
PR 27-MAY-1999; 990S-0136392.  
PR 28-MAY-1999; 990S-0136782.  
PR 01-JUN-1999; 990S-0137222.  
PR 03-JUN-1999; 990S-0137528.  
PR 04-JUN-1999; 990S-0137502.  
PR 07-JUN-1999; 990S-0137724.  
PR 08-JUN-1999; 990S-0138094.  
PR 10-JUN-1999; 990S-0138540.  
PR 10-JUN-1999; 990S-0138847.  
PR 14-JUN-1999; 990S-0139119.  
PR 16-JUN-1999; 990S-0139452.  
PR 16-JUN-1999; 990S-0139453.  
PR 17-JUN-1999; 990S-0139492.  
PR 18-JUN-1999; 990S-0139454.  
PR 18-JUN-1999; 990S-0139455.  
PR 18-JUN-1999; 990S-0139456.  
PR 18-JUN-1999; 990S-0139457.  
PR 18-JUN-1999; 990S-0139458.  
PR 18-JUN-1999; 990S-0139459.  
PR 18-JUN-1999; 990S-0139460.  
PR 18-JUN-1999; 990S-0139461.  
PR 18-JUN-1999; 990S-0139462.  
PR 18-JUN-1999; 990S-0139463.  
PR 18-JUN-1999; 990S-0139750.  
PR 18-JUN-1999; 990S-0139763.  
PR 21-JUN-1999; 990S-0139817.  
  
PR 22-JUN-1999; 990S-0139899.  
PR 23-JUN-1999; 990S-0140353.  
PR 23-JUN-1999; 990S-0140354.  
PR 24-JUN-1999; 990S-0140695.  
PR 28-JUN-1999; 990S-0140823.  
PR 29-JUN-1999; 990S-0140991.  
PR 30-JUN-1999; 990S-0141287.  
PR 01-JUL-1999; 990S-0141842.  
PR 01-JUL-1999; 990S-0142154.  
PR 02-JUL-1999; 990S-0142055.  
PR 06-JUL-1999; 990S-0142390.  
PR 08-JUL-1999; 990S-0142803.  
PR 09-JUL-1999; 990S-0142920.  
PR 12-JUL-1999; 990S-0142977.  
PR 13-JUL-1999; 990S-0143542.  
PR 14-JUL-1999; 990S-0143624.  
PR 15-JUL-1999; 990S-0144005.  
PR 16-JUL-1999; 990S-0144085.  
PR 16-JUL-1999; 990S-0144086.  
PR 19-JUL-1999; 990S-0144325.  
PR 19-JUL-1999; 990S-0144331.  
PR 19-JUL-1999; 990S-0144332.  
PR 19-JUL-1999; 990S-0144333.  
PR 19-JUL-1999; 990S-0144334.  
PR 19-JUL-1999; 990S-0144335.  
PR 20-JUL-1999; 990S-0144352.  
PR 20-JUL-1999; 990S-0144632.  
PR 21-JUL-1999; 990S-0144884.  
PR 21-JUL-1999; 990S-0144884.  
PR 21-JUL-1999; 990S-0145086.  
PR 21-JUL-1999; 990S-0145086.  
PR 22-JUL-1999; 990S-0145085.  
PR 22-JUL-1999; 990S-0145087.  
PR 22-JUL-1999; 990S-0145089.  
PR 22-JUL-1999; 990S-0145192.  
PR 23-JUL-1999; 990S-0145145.  
PR 23-JUL-1999; 990S-0145218.  
PR 26-JUL-1999; 990S-0145224.  
PR 27-JUL-1999; 990S-0145276.  
PR 27-JUL-1999; 990S-0145913.  
PR 27-JUL-1999; 990S-0145918.  
PR 27-JUL-1999; 990S-0145919.  
PR 28-JUL-1999; 990S-0145951.  
PR 02-AUG-1999; 990S-0146386.  
PR 02-AUG-1999; 990S-0146388.  
PR 02-AUG-1999; 990S-0146389.  
PR 03-AUG-1999; 990S-0147038.  
PR 04-AUG-1999; 990S-0147204.  
PR 04-AUG-1999; 990S-0147302.  
PR 05-AUG-1999; 990S-0147192.  
PR 05-AUG-1999; 990S-0147260.  
PR 06-AUG-1999; 990S-0147303.  
PR 06-AUG-1999; 990S-0147416.  
PR 09-AUG-1999; 990S-0147493.  
PR 09-AUG-1999; 990S-0147935.  
PR 10-AUG-1999; 990S-0148171.  
PR 11-AUG-1999; 990S-0148319.  
PR 12-AUG-1999; 990S-0148341.  
PR 13-AUG-1999; 990S-0148565.  
PR 13-AUG-1999; 990S-0148684.  
PR 16-AUG-1999; 990S-0149368.  
PR 17-AUG-1999; 990S-0149175.  
PR 18-AUG-1999; 990S-0149426.  
PR 20-AUG-1999; 990S-0149722.  
PR 20-AUG-1999; 990S-0149723.  
PR 20-AUG-1999; 990S-0149929.  
PR 23-AUG-1999; 990S-0149902.  
PR 23-AUG-1999; 990S-0149930.  
PR 25-AUG-1999; 990S-0150566.  
PR 26-AUG-1999; 990S-0150884.  
PR 27-AUG-1999; 990S-0151065.  
PR 27-AUG-1999; 990S-0151066.  
PR 27-AUG-1999; 990S-0151080.



```
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156536.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 08-OCT-1999; 99US-0158029.
PR 12-OCT-1999; 99US-0158232.
PR 13-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160780.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 9.58; Score 344; DB 21; Length 568;  
Best Local Similarity 27.4%; Pred. No. 9.5e-15;  
Matches 109; Conservative 49; Mismatches 156; Indels 84; Gaps 11;

```
OY 303 IHEPPSLPEPEKTTWKLOORRESSRPVKPKKDPV-----DSQONPA-PEKSSKVEQOLKCC 357
DB 92 IPEYPAVASAPLNNFTSEKNDLGPKKKKOKKNVSGLKRSNPGSDPESEKLLAGMLTFC 151
OY 358 SGLIKENFAKHAAYANPFYKVDVEALGLHDYCDIHKPMDSITIKLEAREYRDAOE 417
DB 152 SOLIVKLMKHK---MAWVFNTLPVDVVGGLHDYHOVYKKRPDLCTVKNLNDKGYVSID 208
OY 418 PGADVRLMFSICYTYNPPDEHVNAMARKLQDVFEKREFAKMPDEPEPVAVSSAPVPPPT 477
DB 209 FADTVRLTFNNAKTYNFKGQDVYFMADKLLDHFDMF-----NPAF 249
OY 478 KVVAPSSSSSSSSSSSSSDSTDDSEERAKRLAELOEQLK-----AVHEQLAALSPQ 532
DB 250 KKPAAQOLKLTGSSSRPEPPKPPFKOKOMONPPNVAAPKKGTEQJSTIAKLLDSVAKPPQ 309
OY 533 QNPKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 592
DB 310 PTLF-----POLVPSRVGSPSPPPPPV-----IQPELPQ-- 341
```

```
OY 593 SKPPPTSESE-----DKCKP-----MSYEEKROLSIDINKLPEEK 629
DB 342 QPPPEQLEIEVEADPDYSEVSKRGKGLPKPKAKDPNRRMLMTMEKSKLGMNLDLPPEK 401
OY 630 LGRVYHIOSREPSLKNKSNPDEIEIDFETLKPSTREL 667
DB 402 LGQLQILKKRNGHLAQDC-DEIETDIADVNETIMEL 438

RESULT 11
AAG41762
ID AAG41762 standard; Protein: 590 AA.
AC AAG41762;
AC AAG41762;
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 52001.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
XX EP1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 28-APR-1999; 99US-0130891.
XX 30-APR-1999; 99US-0131449.
XX 04-MAY-1999; 99US-0132048.
XX 04-MAY-1999; 99US-0132407.
XX 05-MAY-1999; 99US-0132484.
XX 06-MAY-1999; 99US-0132485.
XX 07-MAY-1999; 99US-0132486.
XX 07-MAY-1999; 99US-0132487.
XX 11-MAY-1999; 99US-0132863.
XX 14-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 18-MAY-1999; 99US-0134370.
XX 19-MAY-1999; 99US-0134768.
XX 20-MAY-1999; 99US-0134941.
XX 21-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
```

PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 16-JUN-1999; 99US-0139453.  
 PR 17-JUN-1999; 99US-0139492.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 18-JUN-1999; 99US-0139763.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139899.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 23-JUN-1999; 99US-0140354.  
 PR 24-JUN-1999; 99US-0140695.  
 PR 28-JUN-1999; 99US-0140823.  
 PR 29-JUN-1999; 99US-0140991.  
 PR 30-JUN-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141842.  
 PR 01-JUL-1999; 99US-0142154.  
 PR 02-JUL-1999; 99US-0142055.  
 PR 06-JUL-1999; 99US-0142390.  
 PR 08-JUL-1999; 99US-0142803.  
 PR 09-JUL-1999; 99US-0142920.  
 PR 12-JUL-1999; 99US-0142977.  
 PR 13-JUL-1999; 99US-0143542.  
 PR 14-JUL-1999; 99US-0143624.  
 PR 15-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144085.  
 PR 16-JUL-1999; 99US-0144086.  
 PR 19-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.  
 PR 19-JUL-1999; 99US-0144335.  
 PR 20-JUL-1999; 99US-0144352.  
 PR 20-JUL-1999; 99US-0144632.  
 PR 21-JUL-1999; 99US-0144814.  
 PR 21-JUL-1999; 99US-0144814.  
 PR 21-JUL-1999; 99US-0145086.  
 PR 21-JUL-1999; 99US-0145088.  
 PR 22-JUL-1999; 99US-0145085.  
 PR 22-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145089.  
 PR 22-JUL-1999; 99US-0145192.  
 PR 23-JUL-1999; 99US-0145145.  
 PR 23-JUL-1999; 99US-0145218.  
 PR 23-JUL-1999; 99US-0145224.  
 PR 26-JUL-1999; 99US-0145276.  
 PR 27-JUL-1999; 99US-0145913.  
 PR 27-JUL-1999; 99US-0145918.  
 PR 27-JUL-1999; 99US-0145919.  
 PR 28-JUL-1999; 99US-0145951.  
 PR 02-AUG-1999; 99US-0146386.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 02-AUG-1999; 99US-0146389.  
 PR 03-AUG-1999; 99US-0147038.  
 PR 04-AUG-1999; 99US-0147204.  
 PR 04-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147192.  
 PR 05-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147935.  
 PR 10-AUG-1999; 99US-0148171.

PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148365.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0148175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156599.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161922.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 9.5%; Score 344; DB 21; Length 590;  
 Best Local Similarity 27.4%; Pred. No. 1e-14;  
 Matches 109; Conservative 49; Mismatches 156; Indels 84; Gaps 11;  
 QY 303 IHPPSLPPEPKTKTGQRRESSRPYKPPKKVP----DSQHPA-PEKSSKVSQKCC 357  
 DB 114 IPEVPAVRSAPLNNFTGKERNDLGPKKKKKKVNSGILKRSNPGPDSPESEKTLIAGMLNFC 173



```

XX PF 23-JUL-1997; 97MO-US12877.
XX XX
XX PR 23-JUL-1996; 9605-0022273.
XX XX
XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX XX
XX PI Howard BH, Nakatani Y;
XX XX
XX DR WPI; 1998-120777/11.
XX DR N-PSDB; AAY10093.
XX XX
XX PT New isolated p300/CBP-associated factor, p/CAF - used to develop
XX PT products for modulating transcription, e.g. for treating HIV
XX PT infection or cancers or for promoting muscle differentiation
XX PS Disclosure: Page 81-86; 107pp; English.
XX XX
XX CC This polypeptide sequence comprises CBP, a global transcriptional
XX CC coactivator that is involved in the regulation of various
XX CC DNA-binding transcriptional factors. The invention relates to a
XX CC novel human p300/CBP associated cofactor, p/CAF (see AAM40052), that
XX CC modulates transcription through binding to p300 and CBP (see
XX CC AAM40058). The region (see AAM40060) of CBP that binds to p/CAF
XX CC is claimed. The invention provides methods of screening for
XX CC compounds that inhibit or stimulate the transcription modulating
XX CC and histone acetyltransferase activity of p/CAF and p300/CBP.
XX CC Inhibitors can be used e.g. to inhibit HIV RTF-mediated
XX CC transcription in the treatment of HIV infection. Stimulators can
XX CC be used e.g. to activate tumour suppressor p53 in the treatment of
XX CC cancer or to activate the muscle differentiation factor MyoD to
XX CC promote muscle differentiation. The products can also be used to
XX CC inhibit the cell cycle progression inducing effect of an
XX CC oncoprotein which binds p300/CBP in a subject.
XX XX
SQ Sequence 2441 AA;

Query Match 8.9%; Score 323.5; DB 19; Length 2441;
Best Local Similarity 23.4%; Pred. No. 1.3e-12;
Matches 135; Conservative 72; Mismatches 177; Indels 193; Gaps 27;

OY 9 TRLRLNLPVMDGLTSMSTTQAOAOPANASTNPP--PETSNNPKRKQTNOQLYL 66
DB 670 TRLHKQGIIGN-----QPALPASGAQPVITPAQSVRRPN----- 704
OY 67 LRVYLKTLMKHQPAMFPOQPDVAVKLN--LPDYKIIKTPMDMGTK-----KR 113
DB 705 -----GRLPLRPVNMQVSGQNSP-----NPMISGVNQLQAPMGPRASP 745
OY 114 LNNNYWNAOECT-----QPFNTFTCYIYKRGDVIYMAEA-LKKLELOKT 161
DB 746 MNHSVQMSNMAVPGMAISPRMPORPNMGTGA--NN-----TMAQAPTONOFLPQ- 795
OY 162 NELPTEFETELMIYQAKRGGRKGTGA--KRGVS-----TVPNTQASPPQOTOP 211
DB 796 NQPPS--SSGAMSNVSGKQOPAAQAGVSQGEAALPNPLMLAPQASLPCRPVYQSP 854
OY 212 -QPNPP-----VOATPRP-FPAVTPDLIVQTPVTPVPPPL-----QTPPPVP 254
DB 855 LHTTPPAPASGAMPISLQHTAPGMP-----PQPAAPTPQSPRTVSSGQTPPTPRGVSVPSA 910
OY 255 -----PQPPPPAP-----APQVQSHNPII----- 275
DB 911 AOTQSTPTVHAQAQAVTQPOPTQVPQPSVAPPOSSQOQPTPHOTPRGLTSLQAASID 970
OY 276 --AATPPQVTKKGVKRAKADTTTP-----TTIDPIHEPPSLPPEPKTKLGQRRSS 325
DB 971 NKVPRTSTYTSATSSQAGPQVPMLEKTEVOTDAAEPPEPTEKGPSEMEEDLOQS 1030
OY 326 RPKV-----PPKKDVPDSOHPAPE-----KSSKVSQOLKCCGILKEMPAKR- 368
DB 1031 SOVKBEETDTTEOKSEPMVEYBEKKPEYKVEAKEEENSNDTASQSTSPSQPRKKIKPKPE 1090

```

```

OY 369 -----HAAY-----AMPFTKPDVVEALGLHDYCDIIKHPMDNSTIKLEAREYD 414
DB 1091 LROALMPTLEALYKODESIPFRQPDVDPOLLGIDPDYDIKNNMDISTIRKIDTGQYOE 1150
OY 415 AOEFGADVRLMFSNCKYKPNPDHEVYVAMARKIDVFE 451
DB 1151 PMQYVDVDRIMFNAMILYNKRTSRVYKFCSKLAEVFE 1187

RESULT 14
ID AAY94252
AA AAY94252 standard; protein; 2441 AA.
AC AAY94252;
XX 28-JUL-2000 (first entry)
DE Mouse nuclear CREB Binding Protein, CBP.
KW Mouse; nuclear CREB binding protein; CBP; transcription factor;
KM cyclic AMP; response element; non-insulin dependent diabetes mellitus;
XX GAL4 response element.
OS mus sp.
XX key Location/Qualifiers
XX FT MISC-difference 377
XX FT MISC-difference 377
XX FT MISC-difference 377
XX PN US603583-A.
XX PA (SALK ) SALK INST BIOLOGICAL STUDIES.
XX PI Mounting MR.
XX DR WPI; 2000-375490/32.
XX DR N-PSDB; AAA15554.
XX PT Identifying compounds which disrupt CREB : CBP complex for treating
XX PT diabetes, involves monitoring the expression of a reporter gene in
XX PT response to exposure of the compound, in a modified host cell -
XX PS Disclosure; Column 29-42; 25pp; English.
XX XX
XX CC The present sequence is the mouse nuclear CREB binding protein (CBP).
XX CC CREB is the cyclic AMP (cAMP) response element binding protein. CBP
XX CC cooperates with upstream activators which are involved in transcription
XX CC activation. Signal transduction by cAMP appears to be involved in the
XX CC actiology of non-insulin dependent diabetes mellitus (NIDDM), and so
XX CC inhibition of cAMP signal transduction, which would lead to gene
XX CC transcription inhibition, may be a possible target for therapy. To
XX CC identify possible cAMP activation inhibitors, the expression of a
XX CC reporter construct comprising a GAL4 response element operatively linked
XX CC to a reporter gene can be monitored. Any test compounds which reduce
XX CC reporter gene expression would be a cAMP activation inhibitor.
XX XX
SQ Sequence 2441 AA;

Query Match 8.9%; Score 323.5; DB 21; Length 2441;
Best Local Similarity 23.4%; Pred. No. 1.3e-12;
Matches 135; Conservative 72; Mismatches 177; Indels 193; Gaps 27;

OY 9 TRLRLNLPVMDGLTSMSTTQAOAOPANASTNPP--PETSNNPKRKQTNOQLYL 66
DB 670 TRLHKQGIIGN-----QPALPASGAQPVITPAQSVRRPN----- 704
OY 67 LRVYLKTLMKHQPAMFPOQPDVAVKLN--LPDYKIIKTPMDMGTK-----KR 113

```

Db	705	-----GPIPLPLVNRWYSGQGNSE-----NEMSLGNVQLQADWGPRAAP	745
Oy	114	LENNYVMAQOCI-----QDFNITFTNCYITNKRGGDIIVLMAFA-LEKLEFLQXI	161
Db	746	MNHVSQNSMASVYGMAISPSRMPQPPNMAGTIA-----NM-----TMAQPTQNOFLPO-	795
Oy	162	NELPTEETELMIVQAKGRGGRKETGTA-----KRGVS-----TVENNTQASTPQOTQ	211
Db	796	NOFPS-SSGAMSVNSVGMGQPAAGVSGQGEFGALPLNPLNLAPQAQSLQCPVYQSP	854
Oy	212	-QPNPP-----VQATPHR-FRAVTPDLIVQTPVTVVPPQL--QTPPPVP-----	254
Db	855	LHPPTPPASTAAGMPSLQHTPAQMP-----PQPAAPTOPSPSTVSSGQTPPTPGSVPSA	910
Oy	255	-----POPPPPAP-----APQVQSHPLII-----	275
Db	911	AQOSTPTRYQAAAOAYTPQAPQTPVQPPSVATPQSSQOQTPPHYQTPPGTPLSQAAASD	970
Oy	276	--AATPPVTKRKGVKKRADTTP-----TTIDPIHEPSPLEPPKPTKLGORRESS	325
Db	971	NRVPTSPVTVSAGTSSQAPGPDVPMLEMTKEVQTDAAEPPEPTESKGEPSHMEEDLQGS	1033
Oy	326	RPVK-----PPKRVPPSQGHPAPE-----KSKSVSQQLKCCGSLKEMFAK-	368
Db	1031	SOVKEETDTTEQKSEPMVEEKKRPYEVKVEAKEEENSSNDTASQSTSPQPPRKITFKEP	1090
Oy	369	-----HAAV-----AMPFYKPVQDVEALGLHDYCDIILKHPMDSTIKSLAEAREYD	414
Db	1091	LRLQMLPTEALYQDPESLPFRQPDVDPQLGIPDYFDLYKNPMDLSTIKRLDYGQYE	1150
Oy	415	AOEFGADVRLMFNSCYKYNPPDHEVYVAMARKLDVVE	451
Db	1151	PMQYVDVDRLMFNNAMLYNKRKTSRVYKFCSKLAEPVE	1187
RESULT 15			
ABB44555	ID	ABB44555 standard; Protein; 2441 AA.	
XX	AC	ABB44555;	
XX	DT	25-JAN-2002 (first entry)	
De	DT	Mouse wound healing related polypeptide SEQ ID NO 12.	
XX	KM	Human; mouse; vulnerrary; dermatological; skin disorder; wound healing	
XX	KM	gene therapy.	
XX	OS	Mus musculus.	
XX	PN	CA2325226-A1.	
XX	PD	17-MAY-2001.	
XX	PE	16-NOV-2000; 2000CA-2325226.	
XX	PR	17-NOV-1999; 99DE-1055349.	
XX	PR	17-DEC-1999; 99US-0172511.	
XX	PR	20-JUN-2000; 2000DE-1030149.	
XX	PA	(SWIT-) SWITCH BIOTECH AG.	
XX	PI	Regenbogen J, Wolf E, Goppelt A, Werner S, Halle J;	
XX	DR	WPI; 2001-433142/47.	
PT	PT	Use of novel polypeptide or its variant or nucleic acid encoding the	
PT	PT	polypeptide for diagnosing and/or preventing and/or treating skin	
PT	PT	disorders and/or treatment in wound healing or for identifying active	
PT	PT	substances	
XX	XX	Disclosure; Page 123-132; 265pp; English.	

XX The invention relates to the use of a polypeptide (ABBA4544-ABBA4601, CC ABBA4606-ABBA4623) or its variant or encoding nucleic acid CC (ABBA1990-ABBA1995, ABBA2016-ABBA2032) with vulnery and/or CC dermatological activity for the diagnosis, prevention and treatment of CC skin disorders and treatment in wound healing or for the identification CC of pharmacologically active substances. The nucleic acids are useful in CC gene therapy.

CC Note: The printed sequence listing for this specification was incomplete, CC terminating part way through SEQ ID NO 106. The remaining data was CC obtained from EPO data for an equivalent patent (EP1114862).

XX

SEQ Sequence 2441 AA:

Query Match	8.9%;	Score 323.5;	DB 22;	Length 2441;
Best Local Similarity	23.4%;	Pred. No. 1.3e-12;		
Matches 135;	Conservative 72;	Mismatches 177;	Indels 193;	Gaps 27;

[illegible]

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2003, 13:19:16 ; Search time 32 Seconds  
(without alignments)  
632.593 Million cell updates/sec

Title: US-09-700-590A-22

Perfect score: 3639  
Sequence: 1 MSASGPGTRRLRNLPMVGDG.....ALCHLLFAEKEKTFKRLKM 688

Scoring table: BIOSUM62  
Gapop 10.0 , Gapect 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*

1: /cgn2\_6/prodata/1/1aa/55\_COMB.pep:\*\n2: /cgn2\_6/prodata/1/1aa/55\_COMB.pep:\*\n3: /cgn2\_6/prodata/1/1aa/55\_COMB.pep:\*\n4: /cgn2\_6/prodata/1/1aa/55\_COMB.pep:\*\n5: /cgn2\_6/prodata/1/1aa/55\_COMB.pep:\*\n6: /cgn2\_6/prodata/1/1aa/55\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	330	9.1	65	1	US-08-227-536-5
2	330	9.1	65	5	PCT-US95-04682-5
3	323.5	8.9	2441	1	US-08-194-468-2
4	323.5	8.9	2441	3	US-08-961-739-2
5	323.5	8.9	2441	4	US-09-514-247A-8
6	314	8.6	65	1	US-08-227-536-6
7	314	8.6	65	5	PCT-US95-04682-6
8	308	8.5	2442	4	US-09-514-247A-10
9	305	8.4	2414	1	US-08-227-536-2
10	305	8.4	2414	5	PCT-US95-04682-2
11	209	5.7	331	6	5202236-37
12	209	5.7	865	4	US-09-281-766-19
13	207	5.7	941	4	US-07-757-022B-14
14	207	5.7	1022	4	US-07-757-022B-84
15	207	5.7	1038	4	US-07-757-022B-74
16	207	5.7	1049	4	US-07-757-022B-58
17	207	5.7	1140	4	US-07-757-022B-104
18	207	5.7	1270	4	US-07-757-022B-44
19	207	5.7	1311	4	US-07-757-022B-42
20	207	5.7	1313	4	US-07-757-022B-142
21	207	5.7	1314	4	US-07-757-022B-46
22	207	5.7	1320	4	US-07-757-022B-50
23	207	5.7	1354	4	US-07-757-022B-60
24	207	5.7	1354	4	US-07-757-022B-48
25	207	5.7	1361	4	US-07-757-022B-40
26	207	5.7	1363	4	US-07-757-022B-52
27	207	5.7	1404	4	US-07-757-022B-2

28	207	5.7	1404	4	US-07-757-022B-62	Sequence 62, App1
29	206.5	5.7	1162	2	US-08-728-323A-2	Sequence 2, App1
30	206.5	5.7	1162	4	US-09-298-568-2	Sequence 2, App1
31	205	5.6	238	4	US-09-257-179-80	Sequence 80, App1
32	204	5.6	1872	1	US-08-188-582-14	Sequence 14, App1
33	204	5.6	1872	1	US-08-646-715-14	Sequence 14, App1
34	204	5.6	1893	1	US-08-188-582-11	Sequence 11, App1
35	204	5.6	1893	1	US-08-646-715-11	Sequence 11, App1
36	202.5	5.6	334	6	5202236-3	Patent No. 5202236
37	200.5	5.5	1187	1	US-08-320-559-28	Sequence 28, App1
38	200.5	5.5	1187	3	US-08-545-860D-28	Sequence 28, App1
39	200.5	5.5	1187	5	PCT-US94-04496-28	Sequence 28, App1
40	200.5	5.5	1210	1	US-08-320-559-26	Sequence 26, App1
41	200.5	5.5	1210	3	US-08-545-860D-26	Sequence 26, App1
42	200.5	5.5	1210	5	PCT-US94-04496-26	Sequence 26, App1
43	199.5	5.5	214	1	US-08-217-327-4	Sequence 4, App1
44	194.5	5.3	1185	4	US-09-041-886-23	Sequence 23, App1
45	193.5	5.3	2972	4	US-09-579-181-2	Sequence 2, App1

## ALIGNMENTS

RESULT 1  
US-08-227-536-5  
Sequence 5, Application US/08227536  
Patent No. 5658784  
GENERAL INFORMATION:  
APPLICANT: Eckner, Richard  
APPLICANT: Ewen, Mark  
TITLE OF INVENTION: LIVINGSTON, David  
TITLE OF INVENTION: NUCLEIC ACID, ENCODING TRANSCRIPTION  
NUMBER OF SEQUENCES: FACTOR P300 AND USES OF P300  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes  
STREET: Ten Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/227, 536  
FILING DATE: 14-APR-1994  
CLASSIFICATION: 436  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Ph.D., Kathleen A.  
REGISTRATION NUMBER: 34,380  
REFERENCE/DOCKET NUMBER: DFCI-308XX  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-2290  
TELEFAX: (617) 451-0313  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 65 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHEICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: Internal  
US-08-227-536-5  
Query Match 9.1%; Score 330; DB 1; Length 65;  
Best Local Similarity 87.7%; Pred. 6.7e-17;  
Matches 57; Conservative 4; Mismatches 0; Gaps 0;  
QY 78 QFANPFOQPVDAVKLNDPYIKIKITPMQGIKRLNNYVNAQECIODFNTMTNCTY 137

```

D6      1  QFAMFQFQPDVAYKGLPDIHKIINQPPMDMGIKIRRELENNIYYMAASECMQDFMTMETNCL 60
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY      138 IYNNP 142
        |||||
D6      61 IYNNP 65

```

RESULT 2  
PCT-US95-04682-5  
; Sequence 5, Application PC/TUS95046822  
; GENERAL INFORMATION:  
; ADDITIONAL:

? APPLICANT: NUCLEIC ACID ENCODING TRANSCRIPTION  
 ? TITLE OF INVENTION: FACTOR P300 AND USES OF P300  
 ? TITLE OF INVENTION: FACTOR P300 AND USES OF P300  
 ? NUMBER OF SEQUENCES: 13  
 ? CORRESPONDENCE ADDRESS:  
 ? ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes  
 ? STREET: Ten Post Office Square  
 ? CITY: Boston

```

?
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: PCT/US95/04682

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match	9.18;	Score 330;	DB 5;	Length 65;
Best Local Similarity	87.78;	Pred. No. 6.7e-17;		
Matches 57;	Conservative 4;	Mismatches 4;	Indels	

**OY** 78 QFAMPFGQPDVAVKLNLPDYKIITIKIPDMGTIKRLNNYYWNAQCIDQENTMTNICY 137  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
**DG** 1 QFAMPFGQPDVAVKLGSLPDYKHIIKQPMDMGTIKRRLNNYYWNASCEQMDENTMTNICY 60

RESULT 3  
US-08-194-468-2  
; Sequence 2, Application US/08194468  
; Patent No. 5750336  
; GENERAL INFORMATION:  
; APPLICANT: Montluny, Marc R.

```

1  TITLE OF INVENTION:  ASSAYS FOR THE IDENTIFICATION OF
2  TITLE OF INVENTION:  COMPOUNDS WHICH INHIBIT ACTIVATION OF cAMP AND MITOGENE
3  TITLE OF INVENTION:  RESPONSIVE GENES
4  NUMBER OF SEQUENCES:  3
5  CORRESPONDENCE ADDRESS:
6

```

```

: MEDIUM TYPE: Floppy disk
:
: COMPUTER: IBM PC compatible
:
: OPERATING SYSTEM: PC-DOS/MS-DOS
:
: SOFTWARE: PatentIn Release #1.0, Version #1.25
:
: CURRENT APPLICATION DATA:

```

```

; INFORMATION FOR SEQ ID NO: 2
;
;     SEQUENCE CHARACTERISTICS:
;
;         LENGTH: 2441 amino acids
;         TYPE: amino acid
;         TOPOLOGY: linear
;
;     MOLECULE TYPE: protein
;
US-08-194-468-2

```

Query Match	8.9%	Score 323.5	DB 1	Length 2441
Best Local Similarity	23.4%	Pred. No. 1.5e-14		
Matches 135; Conservative	72;	Mismatches 177;	Indels 193;	Gaps 27

```
QY 9 TRLRLPVMGDGLETSSOMSTTQAOAOOPPAANAATNPP--PETSPPNKKPKROTNOIQL 66
    ||| ::|| ||| || | | : |
Db 670 TRLHKGILGN-----QPALPASGAQPPVTPAGSVRPPN----- 704
```

```

0y 67 LRVLKTLMKHGFAMPEQPVDAVKLN--LPDYKIIKTFMDMGTLK-----KR 113
      | 11::: : : 11: : :
db 705 -----GPLEPLPVNRKMGVSQGMNSF-----NEMSLGNVQLPQADMGPRASP 745

```

```

07 114 DENNIMNAECCI-----QDENIMFTNCTIINKRGDDIVMAEA-DEKLEQKI 161
      :: |:: | | | | | |:: | | | |
Db 746 MNHSQVMNSASVPGMAISPSRMPQRPNNMGTHA---NN-----IMAQAPFNQNFLEP- 795

```

Db 796 NQFBS-SSGAMS VSNYSVGMPAAQAGVSQGEPCALPNPLNMLAPQASQLPCPPVYQSP 854

Db 855 LHP1PPASTAAGMPSLQHP1APGMP-----PPAPAPTPSTPVSSGQTPPTPGSVPSA 910

Db 911 AGTOSTPTVQAAQAQVTPQPTPTPVQPPPSVATPOSSQQQPTPVHTQPPGCTPLSQAAASID 970

Db 971 NRVPPTSTVTSAEISQPGPDVPMLEKTEVQTDDAEEPTESKGEPRSEMMEDLOGS 1030

```

Db 1031 SQVEETDTTEQKSEPMVEVEKRPVKYEAKEEENSSNDTASQSTSPSPQPKKIFRPEE 109
      || :| | : || | : : | :| : :
0v 360 -----HAY-----AMDEYKRVNVEALCTUDVCOTIKYDNDKMCWICVCFADBEVND 414

```

[illegible]



Db 1091 LRQALMPLLEALYKRODEBSLFRQPVDPQDLGIDYEDYVKNPMDLSTIKRKLDTGQYOE 1150  
QY 415 AQEGADVRLMFNSCYKNPPDHEVAMARKLODVE 451  
Db 1151 PMQYVDVRLMFNNAMLYNRKTSRYVKFCSKLAIEVE 1187

RESULT 4  
US-08-961-739-2  
; Sequence 2, Application US/08961739A  
; Patent No. 6063583  
; GENERAL INFORMATION:  
; APPLICANT: Monlunay, Marc R.  
; TITLE OF INVENTION: Methods for Treating Diabetes Mellitus  
; FILE REFERENCE: SALK1650-1  
; CURRENT APPLICATION NUMBER: US/08/961,739A  
; CURRENT FILING DATE: 1997-10-31  
; EARLIER APPLICATION NUMBER: US 194,468  
; EARLIER FILING DATE: 1994-02-10  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 2441  
; TYPE: PRT  
; ORGANISM: Mus  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (1)...(2441)  
; OTHER INFORMATION: Xaa - Any Amino Acid  
US-08-961-739-2

Query Match 8.9%; Score 323.5; DB 3; Length 2441;  
Best Local Similarity 23.4%; Pred. No. 1.5e-14;  
Matches 135; Conservative 72; Mismatches 177; Indels 193; Gaps 27;

QY 9 TRLNLEPYMGDGLTSMSTQAOAOPANASTNPP--PETSINPKPKROTNOLOYL 66  
Db 670 TRLHKQIGLN-----QPALPASGAQPPVLPAAQSVAPN----- 704  
QY 67 LRVYLKILMKHQFAPFQOQPDVAVKLN--LPDYKIIITPKDMGTIK----- 113  
Db 705 -----GRLPLPVNRMQVSGMNSF-----NPMSLGNVQLPQAPKGPRAASP 745  
QY 114 LNNYNNAOECI-----QDFTMTNCTYNNKPPDDIVLMAEA--LEKLFLOKI 161  
Db 746 MNHSVQNSMASVSGMALISPSRMPQPPNMGTGA--NN-----IMAAFTQNOFLPO- 795  
QY 162 NELPTEETELMIYQAKGRGRKETGA--KPGVS-----TVNPTQASTPPQOTP 211  
Db 796 NQFPS--SSGASVSVSGMGPAAQAGVSGQGPALPPLMLAPQASQLPCPPVYQSP 854  
QY 212 -QPNPP-----VQATPRP--FPAVTPDLIVQTPVMTVPPQPL--QTPPPV- 254  
Db 855 LHPTRPPASTAAGMPSLOHPTAPGKTP--POPAAPTQSTPVSQOTPPPTGGSVPSA 910  
QY 255 -----POQPPAP-----APQVOSHPII----- 275  
Db 911 AQTOSTPTVQAAAOQVTPQOTPVQPPSVATPQSSSQOQPPVHTOPRGRLPSLQAAASID 970  
QY 276 --AATPQPVTKKGVKRRKADTTT-----TTIDPIHEPPSLPPEKTKKLQORSS 325  
Db 971 NRVTPTSVTSAETSSQOPGPDVPMLEKTEVQTDAAPEPTESGKGERSSMEEDLQGS 1030  
QY 326 RPVK-----PPKQVPPDSQHPAPE-----KSKVSEOLKCCSGILKEFAK- 368  
Db 1031 SQVKEETDTTEQKSEPEVEKKEVEVVEAKEEENSNDTASOSTSPQPKKIFKPEE 1090  
QY 369 -----HAAY-----AMPFYKPVDEVALGLHDYCDIIKHPDMSTIKSLEAREYD 414  
Db 1091 LRQALMPLLEALYKRODEBSLFRQPVDPQDLGIDYEDYVKNPMDLSTIKRKLDTGQYOE 1150  
QY 415 AQEGADVRLMFNSCYKNPPDHEVAMARKLODVE 451  
:: |||||:: || |||||

Db 1151 PMQYVDVRLMFNNAMLYNRKTSRYVKFCSKLAIEVE 1187

RESULT 5  
US-09-514-247A-8  
; Sequence 8, Application US/09514247A  
; Patent No. 6365361  
; GENERAL INFORMATION:  
; APPLICANT: TANABE SEIYAKU CO. LTD.  
; APPLICANT: TANIGUCHI, Tomoyasu  
; APPLICANT: MIZUKAMI, Junko  
; TITLE OF INVENTION: METHOD FOR IDENTIFYING OR SCREENING AGONIST AND ANTAGONIST TO  
; FILE REFERENCE: TANIGUCHI-6  
; CURRENT APPLICATION NUMBER: US/09/514,247A  
; CURRENT FILING DATE: 2000-02-28  
; PRIOR APPLICATION NUMBER: PCT/JP98/03734  
; PRIOR FILING DATE: 1998-08-24  
; PRIOR APPLICATION NUMBER: JP231084/1997  
; PRIOR FILING DATE: 1997-08-27  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 8  
; LENGTH: 2441  
; TYPE: PRT  
; ORGANISM: mouse  
US-09-514-247A-8

Query Match 8.9%; Score 323.5; DB 4; Length 2441;  
Best Local Similarity 23.4%; Pred. No. 1.5e-14;  
Matches 135; Conservative 72; Mismatches 177; Indels 193; Gaps 27;

QY 9 TRLNLEPYMGDGLTSMSTQAOAOPANASTNPP--PETSINPKPKROTNOLOYL 66  
Db 670 TRLHKQIGLN-----QPALPASGAQPPVLPAAQSVAPN----- 704  
QY 67 LRVYLKILMKHQFAPFQOQPDVAVKLN--LPDYKIIITPKDMGTIK----- 113  
Db 705 -----GRLPLPVNRMQVSGMNSF-----NPMSLGNVQLPQAPKGPRAASP 745  
QY 114 LNNYNNAOECI-----QDFTMTNCTYNNKPPDDIVLMAEA--LEKLFLOKI 161  
Db 746 MNHSVQNSMASVSGMALISPSRMPQPPNMGTGA--NN-----IMAAFTQNOFLPO- 795  
QY 162 NELPTEETELMIYQAKGRGRKETGA--KPGVS-----TVNPTQASTPPQOTP 211  
Db 796 NQFPS--SSGASVSVSGMGPAAQAGVSGQGPALPPLMLAPQASQLPCPPVYQSP 854  
QY 212 -QPNPP-----VQATPRP--FPAVTPDLIVQTPVMTVPPQPL--QTPPPV- 254  
Db 855 LHPTRPPASTAAGMPSLOHPTAPGKTP--POPAAPTQSTPVSQOTPPPTGGSVPSA 910  
QY 255 -----POQPPAP-----APQVOSHPII----- 275  
Db 911 AQTOSTPTVQAAAOQVTPQOTPVQPPSVATPQSSSQOQPPVHTOPRGRLPSLQAAASID 970  
QY 276 --AATPQPVTKKGVKRRKADTTT-----TTIDPIHEPPSLPPEKTKKLQORSS 325  
Db 971 NRVTPTSVTSAETSSQOPGPDVPMLEKTEVQTDAAPEPTESGKGERSSMEEDLQGS 1030  
QY 326 RPVK-----PPKQVPPDSQHPAPE-----KSKVSEOLKCCSGILKEFAK- 368  
Db 1031 SQVKEETDTTEQKSEPEVEKKEVEVVEAKEEENSNDTASOSTSPQPKKIFKPEE 1090  
QY 369 -----HAAY-----AMPFYKPVDEVALGLHDYCDIIKHPDMSTIKSLEAREYD 414  
Db 1091 LRQALMPLLEALYKRODEBSLFRQPVDPQDLGIDYEDYVKNPMDLSTIKRKLDTGQYOE 1150  
QY 415 AQEGADVRLMFNSCYKNPPDHEVAMARKLODVE 451  
Db 1151 PMQYVDVRLMFNNAMLYNRKTSRYVKFCSKLAIEVE 1187  
RESULT 6

```

US-08-227-536-6
; Sequence 6, Application US/08227536
; Patent No. 5658784
; GENERAL INFORMATION:
; APPLICANT: Eckner, Richard
; APPLICANT: Ewen, Mark
; APPLICANT: Livingston, David
; TITLE OF INVENTION: NUCLEIC ACID, ENCODING TRANSCRIPTION
; NUMBER OF INVENTION: FACTOR P300 AND USES OF P300
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes
; STREET: Ten Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/227,536
; FILING DATE: 14-APR-1994
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Ph.D., Kathleen A.
; REGISTRATION NUMBER: 34,380
; REFERENCE/DOCKET NUMBER: DFCI-308XX
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-2290
; TELEFAX: (617) 451-0313
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 65 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: Internal
; US-08-227-536-6

Query Match      8.6%; Score 314; DB 1; Length 65;
Best Local Similarity 84.6%; Pred. No. 9.2e-16;
Matches 55; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 371 AYAMPYKRVVDYDALGLHYCDIIRKPMSTIKSLAREYRDAOEGADYRLMFSNXY 430
DB 1 AYAMPYKRVVDYDALGLHYCDIIRKPMSTIKSLAREYRDAOEGADYRLMFSNXY 60
OY 431 KYNPP 435
DB 61 KYNPP 65

RESULT 7
PCT-US95-04682-6
; Sequence 6, Application PC/TUS9504682
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: NUCLEIC ACID ENCODING TRANSCRIPTION
; NUMBER OF INVENTION: FACTOR P300 AND USES OF P300
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes
; STREET: Ten Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02109
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04682
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/227,536
; FILING DATE: 14-April-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Holliday C. Heine, Ph.D.
; REGISTRATION NUMBER: 34,346
; REFERENCE/DOCKET NUMBER: DFCI-308Xq999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-2290
; TELEFAX: (617) 451-0313
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 65 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: Internal
; PCT-US95-04682-6

Query Match      8.6%; Score 314; DB 5; Length 65;
Best Local Similarity 84.6%; Pred. No. 9.2e-16;
Matches 55; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 371 AYAMPYKRVVDYDALGLHYCDIIRKPMSTIKSLAREYRDAOEGADYRLMFSNXY 430
DB 1 AYAMPYKRVVDYDALGLHYCDIIRKPMSTIKSLAREYRDAOEGADYRLMFSNXY 60
OY 431 KYNPP 435
DB 61 KYNPP 65

RESULT 8
US-09-514-247A-10
; Sequence 10, Application US/09514247A
; Patent No. 6365361
; GENERAL INFORMATION:
; APPLICANT: TANABE SEIYAKU CO. LTD.
; APPLICANT: TANIGUCHI, Junko
; APPLICANT: MIZUKAMI, Junko
; TITLE OF INVENTION: METHOD FOR IDENTIFYING OR SCREENING AGONIST AND ANTAGONIST TO
; FILE REFERENCE: TANIGUCHI-6
; CURRENT APPLICATION NUMBER: US/09/514,247A
; CURRENT FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: PCT/JP98/03734
; PRIOR FILING DATE: 1998-08-24
; PRIOR APPLICATION NUMBER: JP231084/1997
; PRIOR FILING DATE: 1997-08-27
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent version 3.0
; SEQ ID NO 10
; LENGTH: 2442
; TYPE: PRT
; ORGANISM: human
; US-09-514-247A-10

Query Match      8.5%; Score 308; DB 4; Length 2442;
Best Local Similarity 26.1%; Pred. No. 1.9e-13;
Matches 109; Conservative 49; Mismatches 132; Indels 128; Gaps 19;

OY 148 LMAEA-LEKLPLOKINELPTETETIMIVQAKGRGRKRETKTAK---PGVSTV----- 196
;||:| : || : | : : : | : : || :

```

Db 783 MMAAPASQFLPQ--NOFPSSSGAMSV-----GKQPPAQOTGVSQGVFCAALPPLNMLG 837  
QY 197 PNTTQASTPQQTGP-OPNPPVQATPMPFPAVTEDLIYQTPVMTVPPQP-LQTPPPVP 254  
Db 838 PQASQLPCPPPTQSLHPTPPASTA-----AGMPSLQHTTP-PCMTPPQAPAPTPQSTP 891  
QY 255 -----PQPPPPAPAPQVQSHPIIA-----TPQ-----PVKTKKVKKRAKADTTTPTT 299  
Db 892 VSSSGQPTPTPGSVPSATQSTPTVQMAAQAVTPQPPQTPVQPPSVATPQSSQOQPT- 950  
QY 300 IDPIH-EPPSLPPEPKTKTKLGQRESSRPV-----KPKKQVP-----DSOQ 340  
Db 951 --PVHAQPPGPTLSQAAASIDNRVPTBSSVASAETNSQOPQPDVPLEMKTETQAEDETP 1008  
QY 341 HPADK-----SSKVSQOLKCCSGILKE----- 363  
Db 1009 DPGESKGPPEEMMEDLOGASQVKEETDIAEQSEPMVEDEKPEVKEVEKEEESSEN 1068  
QY 364 -----MFAKKHAAY-----AMPFYKPVUVEALGLHDYCDI 393  
Db 1069 GTASQSTSPSPQRRKIKRPEELRQALMPTLEALYRQDESLPFRQPPDPLGIDYEDI 1128  
QY 394 IKHPMDSTIKSKLEAREYRDQAEFGADVRLMFNSCYKNPDPHEVYVAMARKLDVFE 451  
Db 1129 VKNPMDISTIKRKLDGTQYQEPWOYVDDVWLMFNAMLYNKRKTSRYKFKCSKLAEVFE 1186

## RESULT 9

US-08-227-536-2  
Sequence 2, Application US/08227536  
Patent No. 5658784  
GENERAL INFORMATION:  
APPLICANT: Eckner, Richard  
APPLICANT: Ewen, Mark  
APPLICANT: Livingston, David  
TITLE OF INVENTION: NUCLEIC ACID, ENCODING TRANSCRIPTION  
TITLE OF INVENTION: FACTOR P300 AND USES OF P300  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes  
STREET: Ten Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/227,536  
FILING DATE: 14-Apr-1994  
CLASSIFICATION: 436  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Ph.D., Kathleen A.  
REGISTRATION NUMBER: 34,380  
REFERENCE/DOCKET NUMBER: DFCI-308XX  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-2290  
TELEFAX: (617) 451-0313  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2414 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-227-536-2

Query Match 8.4%; Score 305; DB 1; Length 2414;  
Best Local Similarity 22.9%; Pred. No. 3e-13;  
Matches 133; Conservative 65; Mismatches 165; Indels 218; Gaps 24;

QY 9 TRL--RNLPVMDGLETSQMTQAQAPOPANASTNPPPPETS-----NPNKPK-- 57  
Db 650 TRLQKNNMLPNAAGVVPVSNMGPBMGQPP--GMTSNGPLPDDPSIMRGVSNQMMPRIT 707  
QY 58 -----ROTNOLYLVNLTLMKQEFAMFQQPVDAVKLNLPDY 97  
Db 708 POSGLOFGQSMAPPIVPPROTPELOH-----HGQADP----- 742  
QY 98 YKIITPMDKSTIKRLENNYWNQAEIODFTNFTNCYIYNKRGDDIVLMAEALKEFL 157  
Db 743 -GALNPBGVGPBMQPPNQGQFLPQ-----IQFPS----- 772  
QY 158 LQKINELPTEETEMITVQAKR--GGRKKTGTAKPGVSVVPTNTQASTPQQTGPQ- 213  
Db 773 -QGMN-----VTNIPLPASSGQAPVSOQMSSSCPVNSPTMPGQSHIHCPQLPQA 826  
QY 214 ----NPPVQA-----TPHFPAVTEDLIYQTPVMTVPPQQTPTPPVPPQ----- 258  
Db 827 LHQNSPSPVSRTPPHPTP--PSIGAQPPATTI-PAPVPTPAMPBGQSOALHPP 882  
QY 259 -----PPAPAPQPVQSHPIIAATPQPVKTKGKVRADTTPTTIDPIHEPSSLPEP 313  
Db 883 RQTPPTPTQLPQVQPSLPAPASADPQ-----QPSQOSTASV-PTPNAPLLPQP 936  
QY 314 KTKLGQ-----RRSSRPVKPPKK--DVPD----- 337  
Db 937 -ATPLSGPAVSIQGVSNPSTSTEVNSQAIKQPSQEVKMEKMEVDDQPEPADTQPE 995  
QY 338 -----SOQHPPEKSK--VSEOLK 355  
Db 996 DISSEKVEDCKWESTETERSTELKTEKEEDQPSISATQSPAPGSKKIKRPEELR 1055  
QY 356 CCSGLKEMFAKKHAAY-----AMPFYKPVUVEALGLHDYCDI IKHPMDSTIKSKLEAR 410  
Db 1056 -----QALMPTLEALYRQDESLPFRQPPDPLGIDYEDIYKSPMDSTIKRKLDGT 1109  
QY 411 EYRDQAEFGADVRLMFNSCYKNPDPHEVYVAMARKLDVFE 451  
Db 1110 QYEPWOYVDDVWLMFNAMLYNKRKTSRYKFKCSKLAEVFE 1150

## RESULT 10

PCT-US95-04682-2  
Sequence 2, Application PC/TUS9504682  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: NUCLEIC ACID ENCODING TRANSCRIPTION  
TITLE OF INVENTION: FACTOR P300 AND USES OF P300  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes  
STREET: Ten Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/04682  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/227,536  
FILING DATE: 14-April-1994  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Holliday C. Heine, Ph.D.  
REGISTRATION NUMBER: 34,346  
REFERENCE/DOCKET NUMBER: DFCI-308X999



Db 423 LITVGNKMEKORSGKHSFLAESSTGTSTVQALPETDTIST-----PVOERDD 475  
QY 117 NYWNN-----AOECIODFTMTFCYIYNKPGDDIVLMAEALFKLQKINELPTEETEI 171  
Db 476 PYGKKTGSMAPALSGIVALM-----LAANG-----LSATQIRSLIMQSVRTPELSIRV 527  
QY 172 -----MIVQAKRGKRGKREGTAKPGSTVNTQASTPPTQTPQTPQPNP 216  
Db 528 TWGAMPAPKRCIDALALVTPPEGRRPGNPP--SHPPPEASPE-----SSPPRQHHHPHP 581  
QY 217 ---VQATPHPPFAVTDLIQTVMTVPPRQIQTTPPPVPOQPPAPAPQ----- 267  
Db 582 RPNPEASP--PEPS-----PPNMQHPPHPPRPNPPPEASPPESPNNQ 624  
QY 268 -VQSHPLIATPQPVTKKGVKKKADTTPTTIDPIHEPPL-----PPEPTTKLGQ 320  
Db 625 HPHHP-----RPNP-----PGASPPESPNNQHPPHPPRPNP----- 661  
QY 321 RRESSRP-VKPPKQVDSQ-----QHAPPEKSSKVSQKCCGILKEFAKHA 370  
Db 662 --EASPPQSSPPEPQRFPSQWPHTPHFHYHYP----- 693  
QY 371 AYAMPF---KPYDVALGLHDYCDILK--PMDSTIKSL-----EAREYR 413  
Db 694 GYNLPFTYHQSLPYGPG--RDPCPCASHYPADDSPLGSIADPSPQSPYPPSPSK 752  
QY 414 DAOEGADVRLMFNCYKYNPPDEHYVAMARKIQDVFEFRPAKMPD--EPEEPVAVSS 470  
Db 753 PPSPESSPNV-----PSPIRHPSKSLPSAVESPPAPSPQSPHSPEDTSFTK 803  
QY 471 PAVPPPTKVAVP-----PSSSDSSSDSSSDSSTDSEERA 508  
Db 804 PSTPPSPQDPEGRRPESEED--DHKSLSDKSTSHSEGA 843

RESULT 13  
US-07-757-022B-14  
Sequence 14, Application US/07757022B  
Patent No. 6433142  
GENERAL INFORMATION:  
APPLICANT: Gesner, Thomas G.  
APPLICANT: Clark, Stephen C.  
APPLICANT: Turner, Katherine  
APPLICANT: Hewick, Rodney M.  
TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
NUMBER OF SEQUENCES: 143  
CORRESPONDENCE ADDRESS: 143  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/757, 022B  
FILING DATE: 19910910  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/643,502  
FILING DATE: 18-JAN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/546,114  
FILING DATE: 29-JUN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Csert, Luanu  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: GI 5190  
TELEPHONE: (617)876-1170  
TELEFAX: (617)876-5851  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 941 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-757-022B-14

Query Match 5.7%; Score 207; DB 4; Length 941;  
Best Local Similarity 22.0%; Pred. No. 8.6e-07;  
Matches 126; Conservative 47; Mismatches 249; Indels 152; Gaps 23;

QY 161 INELPTEETELMIVQAKRGGRKRE---TGAKPEVST---VNTTQASTPPTQTPOT--- 210  
Db 172 IKSAPTTKEKAPPTTKSAPTTKEKAPPTTKKAPPTTKKAPPTTKKAPPTTKSAPT 231  
QY 211 ---POPNP--PVQATP--HPFPATVTDLIQTP-----VMTVPPQPLQTPP--- 251  
Db 232 TPKEAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAP 291  
QY 252 ---PVPPQPPAPAPQPVQSHPIIATPQPVYTKG--VKRAADTT----- 296  
Db 292 APPTPEKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTK 351  
QY 297 ---PTTIDPIHEP--PSLPEPEKTKRLGRRSSRPVVK---PKQVDSQOHPAPEKSS 348  
Db 352 EPSPTT---TKEPATTTKEKAPT-----TPKKAPPTTKKAPPTTKKAPPTTK 399  
QY 349 KVSQKCCSGILKEMFAKHAAYAMPFYKPVDEALGLHDYCDILKHPMDSTIKSKLE 408  
Db 400 KPAPT-----APKEPAPTTP-----KETAPTTPKKLT 426  
QY 409 AREYRDAOEGADVRLMFNCYKYNPPDEHYVAMARKIQDVFEFRPAKMPDEEP--- 464  
Db 427 PT-----TPKLAFTTPEKAPPTTPEELATTPTEPT-----PTPEPAPT 469  
QY 465 --VAVSSPAVPPPT--KVAPSSSDSSSDSSSDSSTDSEERAQRLAEIQEQLKA 520  
Db 470 PKAAPNTPKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKPKP 529  
QY 521 VHEQLAALSOPOQNKPKKKEKKKKKKHKKRKEEVENKSKAKKEPPKTKKNNSSN 580  
Db 530 APKELAPTT---TKEPSTSDKAPPTPKGTATPTTPREPAPTTPKEKAPPTPK---GT 582  
QY 581 SNVSKKEPAPKSKPPPTT-----SEEDCKCKMSYE-----KRQ 617  
Db 583 APTLKEPAPTTPKPKAPKELAPTTTKGPTSTSDKAPPTPKETAPPTPKKAPPTPK 642  
QY 618 LSLDINKLPGKGLGRVYHIISREPSLKNPNDE 651  
Db 643 PAPTTPPTPTTSEVSTPTTKKEPTTIHKSPDE 676

RESULT 14  
US-07-757-022B-84  
Sequence 84, Application US/07757022B  
Patent No. 6433142  
GENERAL INFORMATION:  
APPLICANT: Gesner, Thomas G.  
APPLICANT: Clark, Stephen C.  
APPLICANT: Turner, Katherine  
APPLICANT: Hewick, Rodney M.  
TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
NUMBER OF SEQUENCES: 143



```

0Y 161 INELTEETELMIYQAGRGGRKE--TGAQGVST-----VPNTQASTPQOTQOT--- 210
Db 237 IKSAPPTPKKEAPPTTYSAPPTPKKEAPPTTKEAPPTPKKEAPPTTKEAPPTTKEAPPTT 296
0Y 211 ---POBNP--PVQANP--HFPFANVTDLIVOP-----VMTVPPOQLQTPP--- 251
Db 297 TPKEBAPPTPKRAPPTTPKEBAPPTTPKEBAPPTTPKEBAPPTTPKEBAPPTTPKEBAPPTPKR 356
0Y 252 ----PVPPPOPPAPAPQYQSHRPLIAATPODVTKKG---VKRKADTT----- 296
Db 357 APPTPKKEAPPTPKKEAPPTTKEBPSPTPKKEBAPPTTKSAPTTTKEBAPPTTKSAPPTPK 416
0Y 297 ---PTIIDPHEP--PSLPREPKTKLQGRRESSRPVR---PKQVDPDSQOHPAPEKS 348
Db 417 EPSPTT---TKERAPPTPKKEBAPT-----TPKRAPPTTPKEBAPPTPKKRAPPTTK 464
0Y 349 KVSBLKQCSGILKEMFAKKAHAAMPFYKRVNVEALGLHDYCDILKHPMDMSTIKSKE 408
Db 465 KRAPT-----APKEBAPTTT-----KETAPTTTPKKLT 491
0Y 409 ARETRDAOEGADYRLMFNSCYKNPDDHEVYAMARKLODFEMKRAKMDDEEP--- 464
Db 492 PT-----TPKILAPPTPKKEAPPTBELAPTTPEEPT-----PTPEEBAPPT 534
0Y 465 ---VVAVSSPAPPEPT--KVYAPRESSSSSSSSSSSSSSDDEEBEERAOGLALQOLKA 520
Db 535 PKAAPPTPKKEAPPTTPKEBAPPTPKRAPPTTPKETAPPTTPKSTAPPTTILKEBAPTPPKR 594
0Y 521 VHEDLAALSOPOOKPKKKEKKKKKKKKKKRKEEVEENKSKAKEPKPKRKNNSN 580
Db 595 APKEAPPTT---TKEPSTISDKNAPPTTPGTATPTTPKEBAPPTTPKEBAPTTTK-----GT 647
0Y 581 SNVSKKEBAPKSKAPPTPE-----SEEBDKCPMSYEE-----KIQ 617
Db 648 APPTTKEBAPTPPKRAPKAPKELAPTTTGPSTISDKNAPPTTPKETAPPTPKBAPTPPK 707
0Y 618 LSLDINKLPGKGLGRVYHIISREPSLAKSNPDE 651
Db 708 PAPPTPEPTPPTTSEVSTPTTTPKEPTTIHKSPOE 741

```

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OK protein - protein search, using sw model

Run on: February 25, 2003, 13:20:47 ; Search time 15 Seconds

(without alignments)  
1425.069 Million cell updates/sec

Title: US-09-700-590A-22

Perfect score: 3639

Sequence: 1 MSASGPGTRRLRLPVMGDG.....ALCHLPLAEKEKFKLRKLM 688

Scoring table: BIOSUM62

Searched: 156504 seqs, 31069816 residues

Total number of hits satisfying chosen parameters: 156504

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA:\*

1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEM\_PUB pep:\*

2: /cgn2\_6/ptodata/2/pubpaa/PC7\_NEM\_PUB pep:\*

3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEM\_PUB pep:\*

4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB pep:\*

5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEM\_PUB pep:\*

6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB pep:\*

7: /cgn2\_6/ptodata/2/pubpaa/PC7US\_PUBCOMB pep:\*

8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB pep:\*

9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEM\_PUB pep:\*

10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB pep:\*

11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEM\_PUB pep:\*

12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB pep:\*

13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEM\_PUB pep:\*

14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	593	16.3	235	10	US-09-764-864-1159
2	564	15.5	140	10	US-09-764-864-1572
3	323.5	8.9	2441	12	US-10-109-886-8
4	308	8.5	2442	12	US-10-109-886-10
5	262.5	7.2	617	10	US-09-864-761-36182
6	220	6.0	128	10	US-09-764-864-1571
7	218.5	6.0	666	10	US-09-791-171-70
8	209	5.7	865	9	US-09-957-995A-19
9	207	5.7	941	12	US-10-124-557-14
10	207	5.7	945	12	US-08-965-272-2
11	207	5.7	945	12	US-10-121-882-2
12	207	5.7	1022	12	US-10-124-557-84
13	207	5.7	1038	12	US-10-124-557-74
14	207	5.7	1049	12	US-10-124-557-58
15	207	5.7	1140	12	US-10-124-557-104
16	207	5.7	1270	12	US-10-124-557-44
17	207	5.7	1311	12	US-10-124-557-42
18	207	5.7	1313	12	US-10-124-557-142
19	207	5.7	1314	12	US-10-124-557-50

20	207	5.7	1320	12	US-10-124-557-46	Sequence 46, Appl
21	207	5.7	1320	12	US-10-124-557-60	Sequence 60, Appl
22	207	5.7	1354	12	US-10-124-557-48	Sequence 48, Appl
23	207	5.7	1361	12	US-10-124-557-40	Sequence 40, Appl
24	207	5.7	1363	12	US-10-124-557-52	Sequence 52, Appl
25	207	5.7	1404	12	US-10-124-557-2	Sequence 2, Appl
26	207	5.7	1404	12	US-10-124-557-62	Sequence 62, Appl
27	205.5	5.6	819	10	US-09-825-144-14	Sequence 14, Appl
28	205	5.6	238	10	US-09-729-835-80	Sequence 80, Appl
29	205	5.6	731	9	US-10-086-464-17	Sequence 17, Appl
30	203.5	5.6	555	10	US-09-764-864-1161	Sequence 1161, Ap
31	203.5	5.6	555	10	US-09-764-864-1574	Sequence 1574, Ap
32	201.5	5.5	5179	9	US-10-025-360-1068	Sequence 1068, Ap
33	201.5	5.5	5179	10	US-09-932-217-1068	Sequence 1068, Ap
34	201.5	5.5	5179	10	US-09-833-263-1068	Sequence 1068, Ap
35	197.5	5.4	580	10	US-09-764-864-1160	Sequence 1160, Ap
36	193	5.3	1125	9	US-09-974-298-114	Sequence 114, App
37	192.5	5.3	160	10	US-09-764-864-1165	Sequence 1165, Ap
38	192	5.3	449	10	US-09-764-864-1163	Sequence 1163, Ap
39	191.5	5.3	332	10	US-09-764-864-1573	Sequence 1573, Ap
40	191	5.2	2783	10	US-09-816-659A-14	Sequence 14, Appl
41	190.5	5.2	1192	10	US-09-758-140-6	Sequence 6, Appl
42	190.5	5.2	1192	10	US-09-972-599A-6	Sequence 6, Appl
43	189.5	5.2	1329	10	US-09-815-242-10112	Sequence 6, Appl
44	187.5	5.2	1192	10	US-09-789-386-2	Sequence 2, Appl
45	187.5	5.2	1192	10	US-09-893-348-23	Sequence 23, Appl

## ALIGNMENTS

RESULT 1

US-09-764-864-1159

Sequence 1159, Application US/09764864

Patent No. US20020132753A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PT23

CURRENT APPLICATION NUMBER: US/09/764,864

Prior application data removed - consult PALM or file wrapper

NUMBER OF SEQ ID NOS: 1792

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 1159

LENGTH: 235

TYPE: PRT

ORGANISM: Homo.sapiens

FEATURE:

NAME/KEY: SITE

LOCATION: (129)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (215)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (221)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-764-864-1159

Query Match

Best Local Similarity 54.1%; Pred. No. 2.3e-23;

Matches 124; Conservative 21; Mismatches 56; Indels 28; Gaps 5;

QY 51 SNPNKPKQTNOLOVLLAVYLKTKKHQFAPFQOPVDAVKIKNTLDYYKIIITTPMDKGTI 110

DB 1 SNPSKPGKRTNLOQYMWQNVATLTKHOFAMFQYOPVDAIKINLPDYHKIIKINPMDKGTI 60

QY 111 KRLNNYVMAQNECIODFNTFTNCYIYNNKGGDVIYMAEAEKLFLOKINELTEFE 170

DB 61 KRLNNYVMAQNECIODFNTFTNCYIYNNKGGDVIYMAEAEKLFLOKINELTEFE 120

QY 171 IMIVQAKGKRGKRTGTAKPGVSTVPNTQASTPPOGTOTPOPNPPVOATPHPPAVTP 230

```
Db 121 ITPCSKG-----XVGSRLREPRRAQVHSHKWPCCPLSDQRPFRACPPSPRRPS 170
Oy 231 DLIVQTPVMTVVP--QP-LQT-----PPVPPQPPPPAPAPQVOSH 271
Db 171 SL-----PPLVQPSLQTSRBSQSPQLPPHLLPHRSSPWSLLRH 209

RESULT 2
US-09-764-864-1572
; Sequence 1572, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1572
; LENGTH: 140
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-864-1572

Query Match 15.5%; Score 564; DB 10; Length 140;
Best Local Similarity 75.4%; Pred. No. 3,9e-22;
Matches 101; Conservative 16; Mismatches 17; Indels 0; Gaps 0;

Oy 51 SNBKRKQTNOLQYLRLVLTLMKHQFAMPQOPVDAYKMLPYYXIKITPMDMGTT 110
Db 1 SNBSKGRKTNOQYQNVVVKTKLMKHQFAMPQOPVDAYKMLPYYXIKITPMDMGTTI 60
Oy 111 KRLNNYWNNAQECIODFNTMFTNCYINRPGDQIVLMAALEKLELQKINELPHEETE 170
Db 61 KRLNNYWNNAQECIODFNTMFTNCYINRPGDQIVLMAALEKLELQKINELPHEETE 120
Oy 171 IMIYQAKGRGRGRK 184
Db 121 ITPCSKGAQVGSR 134

RESULT 3
US-10-109-886-8
; Sequence 8, Application US/10109886
; Patent No. US20020119499A1
; GENERAL INFORMATION:
; APPLICANT: TANABE SEIYAKU CO. LTD.
; APPLICANT: TANIGUCHI, Tomoyasu
; APPLICANT: MIZUKAMI, Junko
; TITLE OF INVENTION: METHOD FOR IDENTIFYING OR SCREENING AGONIST AND
; FILE REFERENCE: TANIGUCHI-6
; CURRENT APPLICATION NUMBER: US/10/109,886
; CURRENT FILING DATE: 2002-04-01
; PRIOR APPLICATION NUMBER: 09/514,247
; PRIOR FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: PCT/JP98/03734
; PRIOR FILING DATE: 1998-08-24
; PRIOR APPLICATION NUMBER: JP231084/1997
; PRIOR FILING DATE: 1997-08-27
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentln version 3.0
; SEQ ID NO 8
; LENGTH: 2441
; TYPE: PRT
; ORGANISM: mouse
US-10-109-886-8

Query Match 8.9%; Score 323.5; DB 12; Length 2441;
Best Local Similarity 23.4%; Pred. No. 6.1e-09;
```

```
Matches 135; Conservative 72; Mismatches 177; Indels 193; Gaps 27;

Oy 9 TRLRLNLPVMDGDLGTSQMSSTTQAQAPANANASTNPP--PETSNNPKRQTNOLQYL 66
Db 670 TRLHKQGIIGN-----QPALPRASGAQPPVTPPAQSVRRPN----- 704
Oy 67 LRVVLKTLMKHOFAMPPOOPVDAVKLN--LPDYKKIKITPMDMGTTK-----KR 113
Db 705 -----GFLPLPVNMQVSGGNSEF-----NMSIGYNQLQAPMGPRASAP 745
Oy 114 LENNYWNNAQECI-----ODFNTMFTNCYINRPGDQIVLMAALEKLELQK 161
Db 746 MNSVQNSMASVPGMAISERMPOPPNNMGTA--NN-----TMAQAPYTONFLPO- 795
Oy 162 NELPTEETELMIYQAKGRGRKETGTA--KPGVS-----TVPNTQASTPPOTGP 211
Db 796 NQFPSSSGMSVNSYGMQAPAGVSGQGEAGALPNPLNLAQASQLPFPYTGQSF 854
Oy 212 -QNPVP-----VQATPRP--PRAVTPDLIVQTPVMTVVPQPL--QTPPVVP----- 254
Db 855 LHPTPPASTAAGMPSLQHPTAGMTP--PQAPAPQPSSTVSSGQTPPTPGSVPSA 910
Oy 255 -----PQPPPPAP-----APQVQSHRPTI----- 275
Db 911 AQTQSTPTVQAQAAQAVTPQPPQTPVQPPSVATPQSSQOQPTPVHTQPTPLSQAAASID 970
Oy 276 --AATPPVTKKGVKRAQDTTP-----TTIDPIHEPPLPEPKTKLQGORRES 325
Db 971 NRPPTSTVTSATSSQOPPPVPMLEKTEYOTDAAEFPTSKKEPSEMAEEDLQGS 1030
Oy 326 RPYK-----PPKQVDPDSQHPAPE-----KSKVSEQLKCCSGILKEMPAKR- 368
Db 1031 SQVKEETDTEQKSEPEMEVEEKKPEVKEAKEEENSNDTASQSTSPQFRKKIKPPE 1090
Oy 369 -----HAAY-----AMPEYKPYDVEALGHDVCDIHKHPMDSTKSKLAEEYND 414
Db 1091 LRQALMPTLEALYRQDPESLPFRQPPVDPOLGIPDYFLVKNPMDSLTKRLKIDTQYOB 1150
Oy 415 AQEFGADVRLMFSNCYKNPDPHEVVAAMARKLDVPE 451
Db 1151 PMQYVDVRLMFMNMLYNNKTSRYVKFKCSKLAEVPE 1187

RESULT 4
US-10-109-886-10
; Sequence 10, Application US/10109886
; Patent No. US20020119499A1
; GENERAL INFORMATION:
; APPLICANT: TANABE SEIYAKU CO. LTD.
; APPLICANT: TANIGUCHI, Tomoyasu
; APPLICANT: MIZUKAMI, Junko
; TITLE OF INVENTION: METHOD FOR IDENTIFYING OR SCREENING AGONIST AND
; FILE REFERENCE: TANIGUCHI-6
; CURRENT APPLICATION NUMBER: US/10/109,886
; CURRENT FILING DATE: 2002-04-01
; PRIOR APPLICATION NUMBER: 09/514,247
; PRIOR FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: PCT/JP98/03734
; PRIOR FILING DATE: 1998-08-24
; PRIOR APPLICATION NUMBER: JP231084/1997
; PRIOR FILING DATE: 1997-08-27
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentln version 3.0
; SEQ ID NO 10
; LENGTH: 2442
; TYPE: PRT
; ORGANISM: human
US-10-109-886-10

Query Match 8.5%; Score 308; DB 12; Length 2442;
Best Local Similarity 26.1%; Pred. No. 3.6e-08;
Matches 109; Conservative 49; Mismatches 132; Indels 128; Gaps 19;
```

QY 148 LMAEA-LEKLEFLAKINELPTEETIMIVQAKGRGRKRTGTAK---PGVSTV----- 196  
Db 783 MMAQAPASOQLPO-NOPSSSSGAMSV-----GMGQPPAOTGVSGQVGAALPPLNLMG 837  
QY 197 PNTQASTPPTOTOP-QNPBPVQATPHPPAVTPBDLVQTPVMTVPPOP-LOTTPPV 254  
Db 838 PQASOLPCBPVTQSPHLHTPPASTA-----ACMPSLQHTTP-PGMPPOAPAPTPSP 891  
QY 255 -----PQPPPPAPAPQPVQSHPIIA-----TPQ---PVTKKGVKKRADDTPPTT 299  
Db 892 VSSSGQPTPTTPGSPVSPATQSTPTVQAAQAVTPOPPQPPVQPPSPATQSSQOQPT- 950  
QY 300 IDPIH-EPPSLPPEPKTKTKGQORRESSRPV-----KPKKDPV-----DSQ 340  
Db 951 --PVHAOPPTPLSQAAASINDRVPPTSSVASAETNSQOPDPVPLEMKTETQADTEP 1008  
QY 341 HPABEK-----SSKVSQDLKCCSGILKE----- 363  
Db 1009 DPGESKGPPESEMEEDLQASQVKEETDIAEQSEPMVEDEKPEVKEVEKEEESSEN 1068  
QY 364 -----MFAKHAAY-----AMPFYKPVDEALGLHDYCDI 393  
Db 1069 CTASQSTSPSPQPKKIKRPEELRQALMPTLEALYRQDESLPFRQPPVDPQLGTPDIFDI 1128  
QY 394 IKHPMDSTIKSLKLEABEYRDQEFQADVRLMFSNCRKYNPDDEHVAMARKLQDVEE 451  
Db 1129 VKNMDLSTIKRKLDTQCYQEPQYVDVWLMFNAMLYNKRKTSRVYKFCSKLAEVFE 1186  
RESULT 5  
US-09-864-761-36182  
; Sequence 36182, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aecmca-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263,6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670

;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 36182  
;; LENGTH: 617  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AC005529.7  
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3  
;; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 0.94  
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3  
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1  
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.3  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.6  
;; OTHER INFORMATION: SWISSPROT HIT: P12036, EVALUATE 2.00e-33  
;; OTHER INFORMATION: EST\_HUMAN HIT: BF445754.1, EVALUATE 3.00e-22  
US-09-864-761-36182  
Query Match 7.2%; Score 262.5; DB 10; Length 617;  
Best Local Similarity 21.9%; Pred. No. 1.8e-06;  
Matches 126; Conservative 96; Mismatches 268; Indels 85; Gaps 20;  
QY 138 IYNKPGDDIVIMALEKFLQKINELPTEETIMIVQAKGRGRKRTGTAKGCVSTVP 197  
Db 43 VKESEKSTVIVEQTE---TQVTEVEETEEKEKEKEEGEGEEBAEG----- 94  
QY 198 NTQASTPPTQTPQP-----NPPVQATPHPPAVTPBDLVQTPVMTVPPQPLQTPPV 253  
Db 95 -EEETKSPABEASPEKAKSPVKEAKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAE 153  
QY 254 PPOPPAPAPAPQVOSHPIIAATPOPKT--KKGVRKADTTPTTI-DPIHEPSLP 310  
Db 154 EAK-SPPAKSPKEEAKSPAEVKSPEKAKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSP 212  
QY 311 PEPTTKLQORRESSRPYKPKKQVDSQHPAPBKSSEVQKCCSGILKEPFAKHA 370  
Db 213 AEAKSPVKEAKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAE 257  
QY 371 AYAMPFYKPVDEALGLHDYCDIIRKHPMDSTIKSLKLEABEYRDQEFQADVRLMFSNCR 430  
Db 258 KSPKAKSPVAEA-----KSP---EKASPVAA-EAKSPKAKSPVKEAKSPAEKSPAE 303  
QY 431 KYNPDDEHVAMARKLQDVEFEMRAKMPDEPEPV-----VAVSSPAVPPTKVVAPP 483  
Db 304 KAKSPVKEAKSPKAKSPVKEA-AKPEKAKSPVKEAKSPKAKSPKAKSPKAKSPKAKSPK 362  
QY 484 SSS 542  
Db 363 AKTAKPEAKSPADKf--PEKAKSPVKEEYKSPKAKSPKAKSPKAKSPKAKSPKAKSPK 419  
QY 543 --KKEKKKHKRKEEVENKSKAKPEPPKTKKNNSSNVSNSVSKKPPAPKSPPTPYE 600  
Db 420 PVKEEKPQEVKVEPPKKAEEKAPATKTEEKD-----SKKEAPKKEAPKVE 472  
QY 601 SEEB--DKCPMSYEERQSLDINKL--PGEKLGAVHIIQGREPSLK-----NSNPE 651  
Db 473 EKKEPAVEKPKRESKVEAKKEAEDEKVKVPTPEKEAPKAVEKEDAKPREKEVAKKREDD 532  
QY 652 IEIDPELKPSTLELGLCHLLEAEKETFKLK 686  
Db 533 AKAK-EPSKPAEKKEA-----APEKKDTKEEK 558  
RESULT 6  
US-09-764-864-1571  
; Sequence 1571, Application US/09764864

```

; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1571
; LENGTH: 128
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (6)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (8)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (14)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1571

```

```

Query Match      6.0%; Score 220; DB 10; Length 128;
Best Local Similarity 37.9%; Pred. No. 5.5e-05;
Matches 44; Conservative 20; Mismatches 50; Indels 2; Gaps 1;

QY 48 PETSNNKRRKQTNOLQYLRLVLTLMKHPAPPOQPPDAVKLMDPYKILITKTPMDM 107
DB 9 PSGRKSKRPEDDQYSLTSLKSLQYKSHQSAWPEPEVK--RTAPGYEYIRPMDL 66
QY 108 GTIKRLNNYNNNAOECIODENTMFTNCYIYKPGDDIYLMAELEKLEKLINE 163
DB 67 KTMSELRKRYVSKRLPMADLRVLTNCKEYNPSEYIKCANILEKFFPSKIKE 122

```

```

RESULT 7
US-09-791-171-70
; Sequence 70, Application US/09791171
; Patent No. US20020094336A1
; GENERAL INFORMATION:
; APPLICANT: ANDERSEN, Peter
; APPLICANT: NIELSEN, Rikke
; APPLICANT: OETTINGER, Thomas
; APPLICANT: RASMUSSEN, Peter Birk
; APPLICANT: ROSENKRANDS, Ida
; APPLICANT: WEDDINGH, Karin
; APPLICANT: FLORIO, Walter
; TITLE OF INVENTION: NUCLEIC ACIDS FRAGMENTS AND POLYPEPTIDE FRAGMENTS
; FILE REFERENCE: DERIVED FROM M. TUBERCULOSIS
; CURRENT APPLICATION NUMBER: US/09/791,171
; CURRENT FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 09/050,739
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 0376/97
; PRIOR FILING DATE: 1997-04-02
; PRIOR APPLICATION NUMBER: 1277/97
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/044,624
; PRIOR FILING DATE: 1997-04-18
; PRIOR APPLICATION NUMBER: 60/070,488
; PRIOR FILING DATE: 1998-01-05
; NUMBER OF SEQ ID NOS: 173
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 70
; LENGTH: 666
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-791-171-70

```

```

Query Match      6.0%; Score 218.5; DB 10; Length 666;
Best Local Similarity 21.8%; Pred. No. 0.00031;
Matches 98; Conservative 33; Mismatches 156; Indels 16; Gaps 15;

QY 190 KRGVSTVPNTQ-----ASTPQOTQPPNPVQATGHPHPAVPDLIVQTPVMTVP 243
DB 44 KPGQGTQPTSDLSRFRFASAPPPPPPPPPPP-----TPMPLAAGHPSPSEPAASKXP 100
QY 244 PQLQTPPPVPPQPPPPA--APQVQSHP-----PIAATPQPVKTKKGVKRAKDTT 296
DB 101 PMPIAGPEPAPKPPPTPMPIAGPEPAPKPPPTPMPIAGPATPTESOLAPRPPTPT 160
QY 297 PPTID-----PIHPPSLPPPKTKTKLGORRESSR 326
DB 161 PTGAPQPPSPAPHPVSHGPHQRTAPAPPMKMMIGEPAPSPR-----SAS 210
QY 327 PVKPPKDVDSQHPAPPEKSKVSEQLCCSGILKEKFAKKAHAAYAMPYKPVDEALG 386
DB 211 PAEPPTP-----PAPQHSRR-----ARRGHRRTDTERNVGKVAATG 246
QY 387 LHDYCDIHKHPMDSTIKSLAREYRDAQFEGADVRLMFSNCKYKNPDDHEVVAARKL 446
DB 247 -----PSTQARLRAEASGAQ-----LABGTE----- 268
QY 447 ODVFENRFAKMPDEPEPVAVVSSPAVPPTKVVAPSSSDSSSDSSDSTDSSEB 506
DB 269 -----PSPAPLQGPSTYLAPTRPAPTE---PPSPSPQNRNG-----R 304
QY 507 RAQRLAEQLKAVHEQLAALSQPOONKPKKKKKKKKKKKKKRKEVEBNKSKAK 566
DB 305 RAER-----RVHDDLAQHAQAQPSITAATGRRRRRAADLDLATORSLRPAK 355
QY 567 EPPPKTKKNNSSNVSKKEPAPMKSKPP 596
DB 356 GKPVKKVKKPK-----PKATKPP 373

```

```

RESULT 8
US-09-957-995A-19
; Sequence 19, Application US/09957995A
; Patent No. US20020165373A1
; GENERAL INFORMATION:
; APPLICANT: Conrad, Patricia A.
; APPLICANT: Barr, Bradd C.
; APPLICANT: Anderson, Mark L.
; APPLICANT: Sverlow, Karen W.
; TITLE OF INVENTION: Recombinant Neospora Antigens and Their Uses
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/957,995A
; FILING DATE: 15-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/215,858
; FILING DATE: 21-MAR-1994
; APPLICATION NUMBER: US 08/327,516
; FILING DATE: 20-OCT-1994
; APPLICATION NUMBER: US 08/645,951
; FILING DATE: 10-MAY-1996

```

APPLICATION NUMBER: US 09/281,766  
FILING DATE: 30-MAR-1999  
ATTORNEY/AGENT INFORMATION:  
NAME: Bastian, Kevin L.  
REGISTRATION NUMBER: 34,774  
REFERENCE/DOCKET NUMBER: 023070-082512US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 865 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 19:  
US-09-957-995A-19

Query Match 5.7%; Score 209; DB 9; Length 865;  
Best Local Similarity 21.8%; Pred. No. 0.0012;  
Matches 127; Conservative 54; Mismatches 214; Indels 188; Gaps 29;

QY 12 RNLVPGGLTSGMTQAOAQPAPN-----ASTNPPRPSNPNKPKRQTNQLO 65  
DB 363 RGADIMHSMGSDASELRQIEOTAOQGIHHAAGNSGRDVTPTPALSTAVEG 422  
QY 66 LIRVYLKTIAMKHO-----FAMFQOPVDVAKLNP--DYKKIKITPMDMTIKRLN 116  
DB 423 LITVGNMMEKQROSKHFLAESSNGTKSVQIALPCTDIYSTI-----PVGERDD 475  
QY 117 NYTN-----AOEIQDENFTNCTIYNKRGDDIVLMAELKFLQKINELPTEETI 171  
DB 476 PYGKMTGTSMAAPALSGIVALM-----LAANPG-----LSATQIRSLIMOSVNTPELSTRV 527  
QY 172 -----MIYQAGRGGRKGTAKPGVSTVNTQTASNPQOTQOPMR 216  
DB 528 TWGAMPAKRLDALVLPREGRRGNP--SHRPEASPE-----SSPRDRQHHPNP 581  
QY 217 ---PVQATPHRPRAVTPDLIVQTVMTVPQRLQTPRPVPPQRPAPAPQ----- 267  
DB 582 RPNPEASR-DEPS-----PRNQHPHNPRRPRRPRESPRRPMQ 624  
QY 268 -VQSHPRPIATPQPVTKKGVKKADTTPTTIDPIHERPSL-----PREPKTKKLGQ 320  
DB 625 HPHNP--RPN-----PGASPPRESSPPMQRHNPRRPNP----- 661  
QY 321 RRESSRP-VKPPKQDVPSO-----QHAPREKSSKVSQOLKCSGILLKEMPAKKA 370  
DB 662 --EASPPQSSPEPPQRPSPQMPHTRHFHYHNP----- 693  
QY 371 AYAMPV---KPYDVEALGLHYCDILKH--PMDMTIKSL-----EAREYR 413  
DB 694 GYNLPYFYHQSPRGYGYG-RDPCQASHNPYPADDSPLGSAVPSPQSVPRPSPSK 752  
QY 414 DAOEGADVRLMFNSCYINPDEHYVAMARKLDVFEKRAKMD---EEEPVVAVSS 470  
DB 753 PSPEGSAPV-----PSPHNPBSRSLPSAVERSPASQPSPHSPSPPTSPTK 803  
QY 471 PAVPPPTKVAVP-----PSSSDSSSDSSSDSDSTDDSEERA 508  
DB 804 PSTPPSPSQDPEGRREPSEED---DHKSLSDKSTSHSSEGA 843

RESULT 9  
US-10-124-557-14

Sequence 14, Application US/10124557  
Patent No. US20020137894A1

GENERAL INFORMATION:

APPLICANT: Turner, Katherine  
Clark, Stephen C.  
Jacobs, Kenneth  
Hewick, Rodney M.  
Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
NUMBER OF SEQUENCES: 143  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/124,557  
FILING DATE: 16-APR-2002  
CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502  
FILING DATE: 18-JAN-1991  
APPLICATION NUMBER: US 07/546,114  
FILING DATE: 29-JUN-1990  
APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Cseitt, Luann  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: GI 5190  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)876-1170  
TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:  
LENGTH: 941 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 14:  
US-10-124-557-14

Query Match 5.7%; Score 207; DB 12; Length 941;  
Best Local Similarity 22.0%; Pred. No. 0.0016;  
Matches 126; Conservative 47; Mismatches 249; Indels 152; Gaps 23;

QY 161 INELPTEETIMIVQAKGRGRKE---GTAKPGVST---VNTTQASTPPOTOT--- 210  
DB 172 IKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPK 231  
QY 211 ---PQNP--PVQATP-HRPVATPDLIVQTP-----VNTVPPQQLQRP--- 251  
DB 222 TPKEPAPTTPKPAFTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPK 291  
QY 252 ---PVPPQPPAPAPQVOSHPPPIAATPQPVYTKG--VKKADTTT----- 296  
DB 292 APPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPK 351  
QY 297 ---PTTIDPIHER-PSLPPPEKTKLGGORRESSRAVVP---PKKVDPSOOHPAPEKSS 348  
DB 352 EPSPTT---TPKEPAPTTPKEPAPT-----TPKKAPATTPKEPAPTTPKEPAPTTPK 399  
QY 349 KVSQQLKCCSGILKEMPAKKAHAAYAMPYKPYDVALGLHYCDILKHHPMDMTIKSL 408  
DB 400 KPAFP-----APKEPAPT-----KETAPTPPKLT 426  
QY 409 AREYDAQEGADVRLMFNSCYINPDEHYVAMARKLDVFEKRAKMPDEEP----- 464  
DB 427 PT-----TPKLIATTPKEPAPTTPPELAPATTPPEPT-----PTTPPEPAPT 469  
QY 465 --VVAVSSPAVPPPI--KVAVPSSSDSSSDSSSDSTDDSEERAQRLAELODLKA 520





```

TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESS: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-APR-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserf, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 74:
SEQUENCE CHARACTERISTICS:
LENGTH: 1038 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 74:
US-10-124-557-74
Query Match          5.7%; Score 207; DB 12; Length 1038;
Best Local Similarity 22.0%, Pred.No. 0.0018;
Matches 126; Conservative 47; Mismatches 249; Indels 152; Gaps 23
QY      161 INELTETEEMIIIVAKGNGRGKRE---TGATKPGVST---VPNTQTASTPOTOT--- 210
       |  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db      237 IKSATPTTKREPAPTITKSAPTTPPKKEPAPTTYTKEPAPTYTKEPAPTYTTSAPT 296
QY      211 ---QPMP--PVQATP--HPPEAVTPDILVTP-----VMTVPPQPLQTPP--- 251
       |  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db      297 TPKEPAPTTPKKPATPTTKEPAPTTPKEPTPTTKEPAPTYTKEPAPTYTKEPAFPAPKPK 356
QY      252 ----VPPQPOPQRPARAQPVQSHPIIATQPVKITKG---VKRKADTTT----- 296
       |  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db      357 APPTPKEPAPPTPKKPAPPTTKESPSPPTPKKEPAPTYTTSAPPTTKEPAPTYTTSAPPTPK 416
QY      297 ---PTTIPRIHEP--SLPPEPKTTKLGORESSRPVKP----PKKDVPDSOOHAPAEKSS 348
       ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db      417 EPSPTT---TKEPATTTYTKEPAFT-----TKKRAPPTPKKEPAPTYTKEPAPTYTTK 464
QY      349 KVSQGLKCCSGTLIKEMFAKHAAYAMPFYKPDVVEALGLHDYCDIIRHMDSTITSKLE 408
       |  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db      465 KPAPT-----APKEPAFTTP-----KETAPTPPKKLIT 491
QY      409 AREYDAOEFGADVLYLMSNCKYKYPDPDEHYVAMARKLDVTEMRAKKPDEPEE---- 464
       |  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db      492 PT-----TPKLIAPPTTPEKPAFTTPEELIAPTTTPEEPT-----PTTPEEPAPTT 534
       |  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
QY      465 -VVAVSSPAVPPPT--KVVAAPPSSSDSSSDSSSDSSTDSEERQRLALEQLKA 520
       |  ::  |  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||

```

```

Db 535 PRAAAPNPKPEAPPTPKPEAPPTPKPEAPPTPKETAPPTPKGTAPPTLKPEAPPTPKPK 594
Oy 521 VHEQJLAISQPOONPKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKSSN 580
Db 595 APEKELAPTT---TKPESTSTSDKPAPTTPKGTAPPTPKPEAPPTPKPEAPTTPK---GT 647
Oy 581 SNVSKPEAPMKSKPPPYE-----SPEEDCKPMSEY-----KRO 617
Db 648 APTTKKEAPPTPKPKAPEKELAPTTTKGPTSTTSKPAPTTPKETAPPTPKPEAPTTPK 707
Oy 618 ISLDINKLGEKIGRVHIIOSREPSLNNSNDE 651
Db 708 PAPPTPEPTPTTSVSTPTTTTKEPTTIHKSPDE 741

RESULT 14
US-10-124-557-58
: Sequence 58, Application US/10124557
: Patent No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Caser, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 1049 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 58:
US-10-124-557-58

Query Match 5.7%; Score 207; DB 12; Length 1049;
Best Local Similarity 22.0%; Pred. No. 0.0018;
Matches 126; Conservative 47; Mismatches 249; Indels 152; Gaps 23;
161 INELPTEETEMIVAKGRGRKE---TGAKPGVST---VPNTQASTPQOT--- 210

```



Db 280 IKSAPPTPKKBPATTTKSAPTTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKSAPT 339  
QY 211 ----PQNP--PVQAP--HPFAVTPDLIVQTP-----VMVVPQPLQTP-- 251  
Db 340 TPKEBPATTTKBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTK 399  
QY 252 ----PVPQOPPPAPAPQVQSHPIIAATPOVKTGK--VKRADTTT----- 296  
Db 400 APPTPKBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKSAPT 459  
QY 297 ----PTIDPHER--PSLPREPKTKLGORRESSPVK--PRKDVDSQOHAPKSS 348  
Db 460 EPSPTT--TKBPATTTKEBPATTT-----TPKKBPATTTKEBPATTTKEBPATTTK 507  
QY 349 KVSQOLKCSGILKEMFAKHAAYAMPYKRVDEALGLHDYCDLIHPRMDSTIKSKLE 408  
Db 508 KPAPT-----APKEPATTP-----KETAPTPPKLT 534  
QY 409 AREYDAQEGADYRLMFSNCKYNPDPHEVAMARKLDVFEKRFKMPDEPEP-- 464  
Db 535 PT-----TPKLAPTTKEBPATTTPELAPTTPEEP--PTTPEBPATTT 577  
QY 465 --VVAVSSPAVPPPT--KVAPSSSSSDSSSDSSSDSSSDSEERAQRLAELQOLKA 520  
Db 578 PKAAPPTPEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTK 637  
QY 521 VHBOLALSOPOQKK 580  
Db 638 APKELATTT--TKEPSTISDKRAPTTPKGTAPTTKEBPATTTKEBPATTTK--GT 690  
QY 581 SNVSKKEPAPKSKPPPTYE-----SEEDKCKPMSYEE-----KRQ 617  
Db 691 APPTKBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTK 750  
QY 618 LSLDINKLPGKLGKRVVHHIOSREPSLKNSPDE 651  
Db 751 PAPTTPPTTSEVSTPTTKEPTTIHNSPDE 784

RESULT 15  
US-10-124-557-104  
; Sequence 104, Application US/10124557  
; Patent No. US20020137894A1  
; GENERAL INFORMATION:  
; APPLICANT: Turner, Katherine  
; Clark, Stephen C.  
; Jacobs, Kenneth  
; Hewick, Rodney M.  
; Gesner, Thomas G.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 Cambridgepark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/124,557  
; FILING DATE: 16-Apr-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION NUMBER:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990

APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Cseri, Luann  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: GI 5190  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)876-1170  
TELEFAX: (617)876-5851  
INFORMATION FOR SEQ ID NO: 104:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1140 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 104:  
US-10-124-557-104

Query Match 5.7%; Score 207; DB 12; Length 1140;  
Best Local Similarity 22.0%; Pred. No. 0.0019;  
Matches 126; Conservative 47; Mismatches 249; Indels 152; Gaps 23;

QY 161 INELPTEETELIMVQAKGRGRKE--TGAKPGVST--VPNTQASTPQOT-- 210  
Db 371 IKSAPTTKEBPATTTKSAPTTPKEBPATTTKEBPATTTKEBPATTTKEBPATTTKSAPT 430  
QY 211 ----PQNP--PVQAP--HPFAVTPDLIVQTP-----VMVVPQPLQTP-- 251  
Db 431 TPKEBPATTTKBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTK 490  
QY 252 ----PVPQOPPPAPAPQVQSHPIIAATPOVKTGK--VKRADTTT----- 296  
Db 491 APPTKBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTK 550  
QY 297 ----PTIDPHER--PSLPREPKTKLGORRESSPVK--PRKDVDSQOHAPKSS 348  
Db 551 EPSPTT--TKBPATTTKEBPATTT-----TPKKBPATTTKEBPATTTKEBPATTTK 598  
QY 349 KVSQOLKCSGILKEMFAKHAAYAMPYKRVDEALGLHDYCDLIHPRMDSTIKSKLE 408  
Db 599 KPAPT-----APKEPATTP-----KETAPTPPKLT 625  
QY 409 AREYDAQEGADYRLMFSNCKYNPDPHEVAMARKLDVFEKRFKMPDEPEP-- 464  
Db 626 PT-----TPKLAPTTKEBPATTTPELAPTTPEEP--PTTPEBPATTT 668  
QY 465 --VVAVSSPAVPPPT--KVAPSSSSSDSSSDSSSDSSSDSEERAQRLAELQOLKA 520  
Db 669 PKAAPPTPEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTK 728  
QY 521 VHBOLALSOPOQKK 580  
Db 729 APKELATTT--TKEPSTISDKRAPTTPKGTAPTTKEBPATTTKEBPATTTK--GT 781  
QY 581 SNVSKKEPAPKSKPPPTYE-----SEEDKCKPMSYEE-----KRQ 617  
Db 782 APPTKBPATTTKBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTK 841  
QY 618 LSLDINKLPGKLGKRVVHHIOSREPSLKNSPDE 651  
Db 842 PAPTTPPTTSEVSTPTTKEPTTIHNSPDE 875

Search completed: February 25, 2003, 13:23:44  
Job time : 21 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2003, 13:18:31 ; Search time 47 Seconds  
(without alignments)  
1407.244 Million cell updates/sec

Title: US-09-700-590A-22

Perfect score: 3639  
Sequence: 1 MSASGSGPTRLRLPVVGDG.....ALCHLLFAEKEKTFKLKLM 688

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR\_73:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR3:\*  
5: PIR4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1869.5	51.4	733	2	RING3 kinase - chi
2	1857	51.0	754	2	female sterile hom
3	1339.5	36.8	2038	2	female sterile hom
4	608.5	16.7	1087	2	hypothetical prote
5	600.5	16.5	1250	2	hypothetical prote
6	570.5	15.7	686	2	bromodomain prote
7	561	15.4	578	2	transcription fact
8	495.5	13.6	638	2	hypothetical prote
9	368.5	10.1	361	2	bromodomain prote
10	339.5	9.3	461	2	hypothetical prote
11	339	9.3	678	2	bromodomain prote
12	323.5	8.9	766	2	hypothetical prote
13	323.5	8.9	2441	2	CREB-binding prote
14	318	8.7	440	2	F2H15.2 protein -
15	310	8.5	2440	2	transcription coac
16	305	8.4	374	2	transcription adap
17	297.5	8.2	1188	2	hypothetical prote
18	285	7.8	369	2	extensin-like prot
19	282	7.7	703	2	hypothetical prote
20	267.5	7.4	1020	1	kinase-like protei
21	267.5	7.4	1020	1	neurofilament trip
22	265.5	7.3	400	2	probable RING3 pro
23	265	7.3	2027	2	hypothetical prote
24	265	7.3	2056	2	protein R10E1.1 f
25	264	7.3	971	2	hypothetical prote
26	260	7.1	1072	1	neurofilament trip
27	259.5	7.1	1274	2	hypothetical prote
28	258	7.1	854	2	neurofilament trip
29	254.5	7.0	7962	2	elastic titin - hu

30	251.5	6.9	1087	1	OPMSH
31	248.5	6.8	1633	2	JC5056
32	243	6.7	606	2	A43427
33	240	6.6	856	2	T16543
34	238	6.5	1879	2	T19481
35	237	6.5	2649	2	T51023
36	234.5	6.4	228	2	S53504
37	233	6.4	769	2	E96613
38	233	6.4	1474	2	F96531
39	232	6.4	1474	2	T20488
40	231	6.3	699	2	T38073
41	231	6.3	1211	2	T42230
42	225	6.2	6642	2	T29757
43	224	6.2	832	1	S71788
44	223.5	6.1	369	2	S20500
45	223.5	6.1	1006	2	T42731

neurofilament trip  
polybrmo 1 - chic  
neurofilament trip  
hypothetical prote  
hypothetical prote  
hypothetical prote  
extensin-like prot  
hypothetical prote  
hypothetical prote  
nucleolar phosphop  
A4 protein - mus  
protein UNC-89 - C  
P/CAF protein - hu  
hydroxyproline-ric  
atrophin-1 related

#### ALIGNMENTS

##### RESULT 1

T28145

C:Species: Gallus gallus (chicken)

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jun-2000

C:Accession: T28145

R:Milne, S.; Kaufman, J.; Beck, S.

submitted to the EMBL Data Library, May 1998

A:Description: DNA sequencing and analysis of the chicken major histocompatibility co

A:Reference number: Z20475

A:Accession: T28145

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1733 <ML>

A:Cross-references: EMBL:AL023516; NID:e1292539; PID:e1292549; PIDN:CAA18965.1

A:Experimental source: clone cbl2

C:Genetics:

A:Gene: RING3

A:Map position: 16

A:Introns: 64/3; 110/3; 158/1; 227/3; 351/3; 394/3; 479/3; 546/2; 650/1; 691/1

C:Superfamily: unassigned bromodomain proteins; bromodomain homology

F:52-109/Domain: bromodomain homology <BR01>

F:323-380/Domain: bromodomain homology <BR02>

Query Match 51.4%; Score 1869.5; DB 2; Length 733;  
Best local similarity 58.3%; Pred. No. 1.8e-76;  
Matches 391; Conservative 69; Mismatches 142; Indels 69; Gaps 14;

QY	26	MSTTQAQAPQPPANASTNPPESTSNPKPKQTNLQYLRLVLTLMKHQFAPRQ	85
DB	1	MASVVALDTLPQ-----ANPPEVSNPKKPKQRYVNOQLYLVKYLKMHQFAPRQ	54
QY	86	PVDAYKMLPDYKIKITKPMQNGTIRKLENNYVWNAQECIQDFMTFTNCYIYVKKPGD	145
DB	55	PVDAYKMLPDYKIKITKPMQNGTIRKLENNYVWNAQECIQDFMTFTNCYIYVKKPGD	114
QY	146	IVLMAEALKEFLQKINLPRFEETIMVQAKGRGRKREGTAPGVSTPNTQASTP	205
DB	115	IVLMAQTEKIFLQKVAQMPPEDEIVVPAK-----NSHKKGASRAAALLAGLTA	165
QY	206	PQOTPPQNPVPVQATPPFPAVTPDLIVQTPVMTVVPQPLQTPPVPPQPPAPAP	265
DB	166	AO-----QVAVSVSVST-AVYTPSPRIATTIYVI-----PHPSVTSAPLL	205
QY	266	QPVOSHPPIAATP--QPVTKKGVKKRADTTPPTIDPI-----HBPSPILPEPKTK	317
DB	206	KSLHSTAVLVLPAPPTPVAKKGVKKRADTTPPTTATATSGSSPSATLAKAKA	265
QY	318	LGQRESSRPVKKPKKVDPSQOHAPPEKSSKVSQQLCCSGIILKEMAKHAAYAMPY	377
DB	266	IPARESGRPKPKKDLPSQOHOT-SKKGKLSQALKYCNGLIKELLSKHAAYAMPY	324

OY	378	KPVDVALGSHVCDIJKHPMDSTIKSKAEARYDQAEAGVYRLMFNSCYKNPNPDH	437
Db	325	KYVDASALGLHDYHEIIKHPMDLSTIKKMHENRYHDAQFAADYKRLMFNSCYKNPNPDH	384
OY	438	EYVAMARKIADYFEMRPAKPDPEE--PVYAVSSPAVPPTKYVAPPPSSD-----	487
Db	385	DYVAMARKIADYFESAKMPDEPODASPSSVSPALLGALSKSSSEESSDEDDEDDE	444
OY	488	-----SSDSSSDSDSTDSEEBERQRLAELOEOLKAVHEOLAALQOPONRKKEK	541
Db	445	DDDEDESSSESSSDSEES--DSEERANRLAELOEOLRAVHEOLAALQSGVSRKKKRE	503
OY	542	DKKEKKKEEHKKEEYBENKSKAKBPPPKTKK-----NNSNSNVSKKEBPAPMKSKPP	597
Db	504	KKKKKSEKHKKGGBDE--SRAQOLRKRAKKGSGGSGGSSKNKKAKAALPAPP	560
OY	598	T--YSESEKCKMSYEEKROLSLDINKLGEKILGRVYHIIQSEPSILKANSNPDETEIDF	656
Db	561	TLYDSESEESKMTYDEKROLSLDINKLPGEKILGRVYHIIQSEPSILDSNPDETEIDF	620
OY	657	ETLKPSTLREL	667
Db	621	ETLKPSTLREL	631

```

RESULT 2
A:56619
female sterile homeotic (fsh) homolog RING3 - human
C:Species: Homo sapiens (man)
C:Date: 21-Jul-1995 #sequence_revision 03-Aug-1995 #text_change 20-Sep-1999
C:Accession: A56619; S18860; S40781
R:Beck, S.; Hanson, I.; Kelly, A.; Pappin, D.J.; Trowsdale, J.
DNA Seq. 2, 203-210, 1992
A:Title: A homologue of the Drosophila female sterile homeotic (fsh) gene in the class I
A:Reference number: A56619; MUID:9232974; PMID:1352711
A:Accession: A56619
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-754 <BRC>
A:Cross-references: EMBL:X62083; NID:g31471; PIDN:AAA68890.1; PID:g31472; EMBL:M80613; N
A>Note: sequence extracted from NCBI backbone (NCBIP:108781)
C:Genetics:
A:Gene: RING3
C:Superfamily: unassigned bromodomain proteins; bromodomain homology
C:Keywords: duplication
E:52-109/Domain: Bromodomain homology <BR01>
E:325-382/Domain: bromodomain homology <BR02>

```

Query Match	51.0%;	Score 1857;	DB 2;	Length 754;
Best Local Similarity	57.7%;	Pred. No. 6.8e-76;		
Matches 390;	Conservative 74;	Mismatches 124;	Indels 88;	Gaps 19;

0Y	44	NPPEPESNPKPKROCNOLUJLRYVLKTLMKROFMPROQVDAVKLRLPYUKTIK	103
Db	13	NPPEPESNPKPKROCNOLUJLRYVLKTLMKROFMPROQVDAVKLRLPYUKTIK	72
0Y	104	PMDSGTIKRLNENNYUWNAOCSIODFTMTMENCYUANKPGDDIVLVAEALKEJLTKINE	163
Db	73	PMDSGTIKRLNENNYUWNAOCSIODFTMTMENCYUANKPGDDIVLVAEALKEJLTKINE	132
0Y	164	LPRETEELMIAQANG---RG---RGKREKGTAKGCSTVNTQASTSPROTQTPQPN	214
Db	133	MPROBEQDELIVTIPKNSHKGAKKLAALDGSVTSIAHQVPASVSSTALUYTPPEL----	186
0Y	215	PRPVQATPRHPRPAVTRDLIVQTRVMTVVPORPLQTRPRVPRBOPRPARAPORVQSHPR	274
Db	187	PTVYLANRPH-----SVISSPLKSLH-----SAGPRLAIVAAAPRA-----	223
0Y	275	IAATPRQVUTKIKGVKRRADTTPTTIDPIHEP--PSLPR--EPKTTKL-GQRESSRPV	328
Db	224	-----QFLAKKKGKKRADDITTPPR--TALLAPGSPAPBPESLEKKAARLPRMKREGPR	277
0Y	329	KPKPKVDPSQOHPAPEKSSKVSBEOLKCGSGLIKEMFAKKNHAUVAMPFYKPVNVEALGLH	388

[illegible]

RESULT 3  
 A43742 female sterile homeotic protein, 205K - fruit fly (*Drosophila melanogaster*)  
 N:Alternate names: membrane protein fish, 205K  
 N:Contains: female sterile homeotic protein, 110K  
 C:Species: *Drosophila melanogaster*  
 C:Date: 03-Mar-1993 #sequence\_revision 03-Mar-1993 #text\_change 20-Sep-1999  
 C:Accession: A43742; B43742  
 R:Haynes, S.R.; Mocer, B.A.; Bhatia-Dey, N.; Dawid, I.B.  
 Dev. Biol. 134, 246-257, 1989  
 A:Title: The *Drosophila* fish locus, a maternal effect homeotic gene, encodes apparent  
 A:Reference number: A43742; MUID:89276730; PMID:2567251  
 A:Accession: A43742  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-2038 <MAV>  
 A:Cross-references: EMBL:M23221; NID:g157452; PIDN:AAA28540.1; PID:g157453  
 A:Accession: B43742  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-1106 <BA2>  
 A:Cross-references: EMBL:M23222  
 C:Genetics:  
 A:Gene: fish  
 A:Cross-references: FlyBase:FBgn0004656  
 C:Superfamily: unassigned bromodomain proteins; bromodomain homology  
 C:Keywords: alternative splicing; transmembrane protein  
 F:1-2038/Product: female sterile homeotic protein, 205K #status predicted <MA2>  
 F:1-1106/Product: female sterile homeotic protein, 110K #status predicted <MA7>  
 F:59-116/Domain: bromodomain homology <BR01>  
 F:503-560/Domain: bromodomain homology <BR02>

Query Match	36.8%;	Score 1339.5;	DB 2;	Length 2038;
Best Local Similarity	31.9%;	Pred. No. 2e-52;		
Matches 334;	Conservative 91;	Mismatches 173;	Indels 449;	Gaps 21

Qy	31	AAQAPNAASTJNP-----PPETSNPNKRCROTNOLOYLRLVLTLLTKHNDQAPFQO	85
Db	2	SSSEPPPEPPEPPEPVGSIQPPVPIPAEREGCRNTNOLYLKIKRYVMYIKHHNSPFFQO	61
Qy	86	PYDAVKLNLPPYYKIITPMDMGSTIKRLLENNTYMAAECIODENFMFTNCYIYNKFGDD	145
Db	62	PVDAKKLNLPYKHIIQPPMDMGSTIKRLLENNTYMAAECIODENFMFTNCYIYNKFGDD	121
Qy	146	IVLAALAELEKFLQIKINELPPEETEEIIMIVAAKGRRGRKGTGAK-----PGVSTVPEPTT	200













Db 652 IDSDVDETLWEL 663

RESULT 13  
S39161  
CREB-binding protein - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 25-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 02-Aug-2002  
C:Accession: S39161  
R:Chrivia, J.C.; Kwok, R.P.S.; Lamb, N.; Haglwaara, M.; Montminy, M.R.; Goodman, R.H.  
Nature 365, 855-859, 1993  
A:Title: Phosphorylated CREB binds specifically to the nuclear protein CBP.  
A:Reference number: S39161; MUID:94019866; PMID:8413673  
A:Accession: S39161  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-2441 <CHR>  
A:Cross-references: GB:S66385; NID:9435854; PIDN:AAB28651.1; PID:9435855  
C:Superfamily: Transcription coactivator CREB-binding protein; bromodomain homology  
F:1112-1169/Domain: bromodomain homology <BRO>

Query Match 8.9%; Score 323.5; DB 2; Length 2441;  
Best Local Similarity 23.4%; Pred. No. 4.2e-07;  
Matches 135; Conservative 72; Mismatches 177; Indels 193; Gaps 27;

QY 9 TRLRLPVNGDGLTSSMTTQAQAPQANAASTNPP--PETSNNPKKROTNOLQYL 66  
Db 670 TRLHKGILGN-----QPALPASGAQPVIPPAQSVPRPN----- 704  
QY 67 LRVYLKTLKHQFAMPPOQPVDAVKLN--LPDYKILIKTPMDKGLK-----KR 113  
Db 705 -----GLPLRPVNMQVSGQNSF-----NPMSLGNQLDQAPMGPRASP 745  
QY 114 LNNYVWNAOECI-----ODFNMTTCYIYKRGDDIVLAAE--LEKLFLOKI 161  
Db 746 MNHSVOMNSASYPGMAISPRMPORPNMAGTIA--NN-----TMAQAPTONQIFPQ- 795  
QY 162 NELPTEETELMIVQAKRGGRKGTGA--KPGVS-----TVPTTQASTPPOPTP 211  
Db 796 NQPPS--SSGAMSVNAGMGPAPACVSGQEPGALPNLMLAPQASQLPCPPVYQSP 854  
QY 212 -QGNPP--VQATPHH--FPAVTPDLVQFPMVTVPRQL--QTPPPPP----- 254  
Db 855 LHPETPPASTAGMPSLDQHTAPGMP--PQPAAPTOSTVSSGQPTTPPGSVPSA 910  
QY 255 -----PQPPPPAP-----APQVQSHPII----- 275  
Db 911 AQOSTPTVQAAQAQVTPQPTQVQPPSVATPQSSQOQPTVHTQPRGTRLSQAAASID 970  
QY 276 --AATPQPVKTKKVKRKADTTTP-----TTIDPHEPRLPPRPKTKLQRRSS 325  
Db 971 NRPPTPTVTSASPTSSQGPDPVMLEMKTEVQTDAEPRTSGSKRESEMEEDLOGS 1030  
QY 326 RPYK-----PPKDDVDSQHPAP-----KSSVSVQLKCCSILLKEMFAK- 368  
Db 1031 SOVKETDTTEQSESPMEVEKKPEVKEAKEEENSSNNTAQSSTSPSPRRKIKRPEE 1090  
QY 369 -----HAAY-----AMPYKPVDEALGLHDYCDIIKHPMDSTIKSLAEAYRD 414  
Db 1091 LKQALMPTLEALYROPDESIFPQPPDLPOLGIPDYFDIYKPNMDLSTIKRLDTGOYOE 1150  
QY 415 AOEFGADVRLMFSNCKYKPNPDHEVVAAMARKLDVFE 451  
Db 1151 PMQYVDVRLMFMNMLYNNKTSRYVKFKCSKLAEFE 1187

RESULT 14  
H86312  
P2H15.2 protein - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Nov-2001  
C:Accession: H86312

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon  
Chin, C.W.; Chung, M.K.; Comn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,  
ansen, N.F.; Hughes, B.; Hultzer, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: H86312  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-440 <STO>  
A:Cross-references: GB:AEO05172; NID:99665057; PIDN:AAF97259.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1

Query Match 8.7%; Score 318; DB 2; Length 440;  
Best Local Similarity 27.0%; Pred. No. 1.4e-07;  
Matches 93; Conservative 46; Mismatches 83; Indels 122; Gaps 9;

QY 355 KCCSGIIEKFAKKHAYKMPFYKPPDVLEALGLHDYCDIIKHPMDSTIKSLAEAYRD 414  
Db 134 KNCNSLTITKLM--KHS--AWFVNPVDKAGLGLDHYHNHYKEPMDGLTYVTKLGKSLYS 190  
QY 415 AOEFGADVRLMFSNCKYKPNPDHEVVAAMARKLDVFEEMFAKRPDPE----- 462  
Db 191 PLDFADVRLTFNNALILYNPIGHDYTRFALLNMEDKKVSTEMOYDNLHKKFKPTPDI 250  
QY 463 -----EPVAV-----SSPAPPTKYVAPSSSSSSSSSSSSSSSSSSSS 503  
Db 251 EFPAPAPSIAPVIEPLPAIVSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 293  
QY 504 EERRAQRALQOLAVHQLAALSOPOQNKKKKKKKKKKKKKKKKKKKKKKKKKKK 563  
Db 294 -----RTWERESMTIPVEPAVITAP----- 316  
QY 564 KAKEPPPKKKKKNNSSNSVSKKEPAPMSKPPPYESEEDCKPMSEKROLSDIN 623  
Db 317 KAFE-----EAPVNN-----RULTEKRLSDELO 343  
QY 624 KLPEKLGAVYHIIQSREPSLKNNDPEIETIDETIKPSTLREL 667  
Db 344 DLPYDKLETVVQIIKKSNPESQKD--DEIETIDSDIDINTIMEL 386

RESULT 15  
S39162  
transcription coactivator CREB-binding protein - human  
N:Alternate names: CBP; RSTs; Rubinstein-Taybi syndrome (Rts) protein  
C:Species: Homo sapiens (man)  
C:Date: 07-Oct-1994 #sequence\_revision 17-Nov-1995 #text\_change 02-Aug-2002  
C:Accession: S39162; S60345; I58096  
R:Chrivia, J.C.; Kwok, R.P.S.; Lamb, N.; Haglwaara, M.; Montminy, M.R.; Goodman, R.H.  
Nature 365, 855-859, 1993  
A:Title: Phosphorylated CREB binds specifically to the nuclear protein CBP.  
A:Reference number: S39161; MUID:94019866; PMID:8413673  
A:Accession: S39162  
A:Molecule type: mRNA  
A:Residues: 1-2440 <CHR>  
A:Note: differences with the mouse sequence are shown  
R:Lundblad, J.R.; Kwok, R.P.S.; Laurence, M.E.; Harter, M.L.; Goodman, R.H.  
Nature 374, 85-88, 1995  
A:Title: Adenoviral ELA-associated protein p300 as a functional homologue of the tran  
A:Reference number: S60344; MUID:95174889; PMID:7870179  
A:Accession: S60345  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 'S', 574-670, 'T', 672-681 <LUN>  
R:Pettili, F.; Giles, R.H.; Dauwerse, H.G.; Saris, J.J.; Hennekam, R.C.; Masuno, M.; T  
Nature 376, 348-351, 1995

A:Title: Rubinstein-Taybi syndrome caused by mutations in the transcriptional co-activator  
A:Reference number: 158096; MIM:158096; PMID:7630403  
A:Accession: 158096  
A>Status: translation not shown; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 352-356 <P>  
A:Cross-references: GB:089354; NID:91888537; GB:S78936; NID:91041931  
A:Note: this translation is not annotated in GenBank entry S78936, release 112.0  
A:Note: this sequence with a termination mutation is from a patient with Rubinstein-Taybi  
C:Genetics:  
A:Gene: GDB:CREBBP; RTS; CBP; RSTS  
A:Cross-references: GDB:437159; OMIM:180849; OMIM:600140  
A:Map position: 16p13.3-16p13.3  
A:Note: defects in this gene may result in Rubinstein-Taybi syndrome  
C:Superfamily: transcription coactivator CREB-binding protein; bromodomain homology  
C:Keywords: phosphoprotein; transcription; zinc finger  
F:462-661/Domain: CREB binding #status predicted <CB>  
F:111-1168/Domain: bromodomain homology <BRO>  
F:1283-1311/Region: zinc finger CCCC motif  
F:1707-1733/Region: zinc finger CCCC motif  
F:78,381,745,1172,1533,2063,2354/binding site: phosphate (Ser) (covalent) (by protein kinase A) #status predicted  
F:1771/binding site: phosphate (Ser) (covalent) (by protein kinase A) #status predicted

Query Match 8.54; Score 310; DB 2; Length 2440;  
Best Local Similarity 25.28; Pred. No. 1.7e-06;  
Matches 105; Conservative 54; Mismatches 133; Indels 124; Gaps 17;

QY 148 LMAEA-LEKLEFLOKINELPFEETIMIVQAKRGKREKTAK--PGVSTV----- 196  
DB 783 MMAQAPASQOFLPQ-NQFPSSGAMSV---GKQPPAQGTSGQVPGALPPLNMLG 837  
QY 197 PNTQASTPQOTQTP-OPNPPVQATPHPPAVTPDLIVQTVMYVPPQPL----- 247  
DB 838 POASQLPCPVPVQGLPHTPPASTA-----AGMPSLOHTTP-DGMPDPQAPAPTPQSTP 891  
QY 248 -----QTPPPVP-----POP-----QPPAPAPVQSHP-P 273  
DB 892 VSSSGQTPPTPGSVPSATQSTPTVQMAAQAVTPQPPVQPPSVATPQSSQOQPTP 951  
QY 274 IIAATPQPVKTKGKVKRRKADTTPTTI-----DPIHEPP 307  
DB 952 VHAQPTPTPLSQAAASIDNRPTSTYASATNSQOPGPDVPLVLEMKETQAEDETPDPG 1011  
QY 308 SLPEPKTKLGGQRRESSRPVKKDVPDSQHP-----APEKSSKVSQQLKCCSG-- 359  
DB 1012 ESKGEPRSEMMEDLQASQVK-ETDIAEQKSEPMVEDEKPPVKVEKKEEESSSNGT 1070  
QY 360 -----IIEKFAKKHAY-----AMPFYKPYDVEALGLHDCDIK 395  
DB 1071 ASQSTSPSQPRKIKFKEEELRQALMPTLEALYRQDPESLPRQPPVDPQLGIPDYFDIVK 1130  
QY 396 HPMDSTIKSKLEAREYDAQEFQADYILFMSNCKYKNPDHEVYAMARKLQDVEE 451  
DB 1131 NPMDLSTIKRKLDTGQVQEPQVYDDVWLMFNMMALYNRKTSRYKFKCSKLAIFYE 1186

Search completed: February 25, 2003, 13:22:44  
Job time : 55 secs

**THIS PAGE BLANK (USPTO)**



```
FT DOMAIN 974 986 POLY-PRO.
FT DOMAIN 1011 1014 POLY-PRO.
FT DOMAIN 1028 1033 POLY-PRO.
FT DOMAIN 1283 1303 POLY-GLN.
FT DOMAIN 1301 1308 POLY-ALA.
FT DOMAIN 1335 1338 POLY-ARG.
FT CONFLICT 720 721 EM -> GP (IN REF. 2).
SQ SEQUENCE 1362 AA; 152219 MW; D52FEFICF9960907 CRC64;

Query Match
Best Local Similarity 97.0%; Score 3531.5; DB 1; Length 1362;
Matches 672; Conservative 0; Mismatches 5; Indels 9; Gaps 1;

Qy 1 MSAESGPGTRLNLPVMDGDLTSMSTTQAQAQPPANAASSTNPPPPETSNPNKPKROT 60
Db 1 MSAESGPGTRLNLPVMDGDLTSMSTTQAQAQPPANAASSTNPPPPETSNPNKPKROT 60
Qy 61 NOLQYLLRVLLTKLWKHGFANFPQPDVAVKLNLDYKIKTPMDMTIKKRLNNYVW 120
Db 61 NOLQYLLRVLLTKLWKHGFANFPQPDVAVKLNLDYKIKTPMDMTIKKRLNNYVW 120
Qy 121 NAOECIQDFNTFTNICYLYNKPDDIVLMAEALFKLQKINELPTEETIMIVQAKRG 180
Db 121 NAOECIQDFNTFTNICYLYNKPDDIVLMAEALFKLQKINELPTEETIMIVQAKRG 180
Qy 181 RGRKETGAKGVSVPNTQASTPQOTQPPNPPVQATPHPPVATPDIVQTPVMT 240
Db 181 RGRKETGAKGVSVPNTQASTPQOTQPPNPPVQATPHPPVATPDIVQTPVMT 240
Qy 241 VVPQPLQTPPPVPPQPPAPAPQVSHPPPIIAATPPQVTKKGRKADTTPTTI 300
Db 241 VVPQPLQTPPPVPPQPPAPAPQVSHPPPIIAATPPQVTKKGRKADTTPTTI 300
Qy 301 DPHIEPPLPEPKTKLQGRRESSRPVKKKDVDPDQHPAPKSKVSEQLKCCSGI 360
Db 301 DPHIEPPLPEPKTKLQGRRESSRPVKKKDVDPDQHPAPKSKVSEQLKCCSGI 360
Qy 361 LKEMFAKHAAYAWPYFVDPVDEALGLHDYCDIHKHPMDSTIKSLBAREYRDAEFGA 420
Db 361 LKEMFAKHAAYAWPYFVDPVDEALGLHDYCDIHKHPMDSTIKSLBAREYRDAEFGA 420
Qy 421 DVRLMFSNICYKNPDHVEVAMARKLDQVEMRFAKMDPEEPVAVSSPAVPPTKV 480
Db 421 DVRLMFSNICYKNPDHVEVAMARKLDQVEMRFAKMDPEEPVAVSSPAVPPTKV 480
Qy 481 APPSSSSSSSSSDSDSDTDEEERARLAEQELKAVHEQLAALSQPOQNKPKKE 540
Db 481 APPSSSSSSSSSDSDSDTDEEERARLAEQELKAVHEQLAALSQPOQNKPKKE 540
Qy 541 KDKKKKKKKKKKEVEENKSKAKPPPKTKKNNSSNSVSKKEPAPMKSKPPPTYE 600
Db 541 KDKKKKKKKKKKEVEENKSKAKPPPKTKKNNSSNSVSKKEPAPMKSKPPPTYE 600
Qy 601 SEEDKCKPMSVEERKQLSLDINKLPGLKGRVHVHIQSRPSLKNPNDETEIDFETLK 660
Db 601 SEEDKCKPMSVEERKQLSLDINKLPGLKGRVHVHIQSRPSLKNPNDETEIDFETLK 660
Qy 661 PSTRELKALCHLLFAEEKETFKLRK 686
Db 661 PSTRELKALCHLLFAEEKETFKLRK 686
```

## RESULT 2

```
BRD3_HUMAN
ID BRD3_HUMAN STANDARD; PRT; 726 AA.
AC Q15059; O92645;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Bromodomain-containing protein 3 (RING3-like protein).
GN BRD3 OR RING3L OR KIAA0043.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=96051398; PubMed=7584044;
RA Nomura N., Nagase T., Miyajima N., Sazuka T., Tanaka A., Sato S.,
  Seki N., Kawabayashi Y., Ishikawa K.-I., Tabata S.;
  "Prediction of the coding sequences of unidentified human genes. II.
  The coding sequences of 40 new genes (K1AA0041-K1AA0080) deduced by
  analysis of cDNA clones from human cell line KG-1.";
  DNA Res. 1:223-229(1994).
RN [2]
RP SEQUENCE OF 363-726 FROM N.A.
RX MEDLINE=98038990; PubMed=9373153;
RA Thorpe K.L., Gorman P., Thomas C., Sheer D., Trowsdale J., Beck S.;
  "Chromosomal localization, gene structure and transcription pattern of
  the ORFX gene, a homologue of the MHC-linked RING3 gene.";
  Gene 200:177-183(1997).
RL CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- TISSUE SPECIFICITY: UBIQUITOUS.
CC -1- SIMILARITY: CONTAINS 2 BROMODOMAINS
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC
CC EMBL; D26362; BAA05393.1; -
CC EMBL; Z81330; CAB03630.1; -
CC HSSP; Q92831; 1B91.
CC Genew; HGNC:1104; BRD3.
CC MIM; 601541; -
CC InterPro; IPR001487; Bromodomain.
CC Pfam; PF00439; bromodomain; 2.
CC PRINTS; PR00503; BROMODOMAIN.
CC SMART; SM00297; BROMO; 2.
CC PROSITE; PS00633; BROMODOMAIN_1; 2.
CC PROSITE; PS00633; BROMODOMAIN_2; 2.
CC Bromodomain; Repeat; Nuclear protein.
CC DOMAIN 56 115 BROMODOMAIN 1.
CC DOMAIN 326 398 BROMODOMAIN 2.
CC FT DOMAIN 487 555 LYS-RICH.
CC FT DOMAIN 676 725 SER-RICH.
CC FT CONFLICT 465 466 EL -> DV (IN REF. 2).
CC SEQUENCE 726 AA; 79541 MW; 64F526FC3C1033AA CRC64;

Query Match
Best Local Similarity 52.5%; Score 1910.5; DB 1; Length 726;
Matches 396; Conservative 66; Mismatches 144; Indels 55; Gaps 15;

Qy 27 STTQAQAPQAPANAASSTNPPPPETSNPNKPKROTQLOQLLRVLLTKLWKHGFANFPQ 86
Db 4 ATTAVAP-GIPATPGVNPVPPPEVSNPSKPKRTNQLQYMQNVVYVVKLWKHGFANFPQ 62
Qy 87 VDAVKLNLDYKIKTPMDMTIKKRLNNYVNAQECIQDFNTFTNICYLYNKPDDI 146
Db 63 VDAIKLNLDYKIKTPMDMTIKKRLNNYVNAQECIQDFNTFTNICYLYNKPDDI 122
Qy 147 VLMAEALFKLQKINELPTEETIMIVQAKGRGR-GRKETGAKGVSVPNTQASTPQOTQPPNPPVQATPHPPVATPDIVQTPVMTIKKRLNNYVW 202
Db 123 VLMAEALFKLQKINELPTEETIMIVQAKGRGR-GRKETGAKGVSVPNTQASTPQOTQPPNPPVQATPHPPVATPDIVQTPVMTIKKRLNNYVW 180
Qy 203 STTQOTQTP-QPNPPVQATPHPPVATPDIVQTPVMTIKKRLNNYVW 261
Db 181 -----TPFQSVPPVTSQTP-----VIAATFPVTITAN---VTSVPVPPAAAPP 221
Qy 262 APAQVQSHPPITAAATPPQVTKKGRKADTTPTT-----TIDPIHEPPSLPPPKTK 317
Db 222 -----PATPIVPPVPTP-PVKKKGRKADTTPTTSAITASRSSESPPL-SDPKQAK 274
```

```
QY 318 LQORRES-SRPVKKDDVPDSQOHPAPEKSKSVSEQLKCCSGILKEMFAKKHAAAYAWPF 376
: ||||| :||:|||||:| :| :||:| :||:| :||:| :||:| :||:|
Db 275 VVARRESGGRPKPPKDLKEDGVEYPOHAGKKGLKSEHLRYCDSTLREMLSKHAAAYAWPF 334
: ||||| :||:|||||:| :| :||:| :||:| :||:| :||:| :||:|
QY 377 YKPYDVEALGLHDYCDLIKHPMDMSTTKSLKLEAREYDAQOFGADVRLMFSCYKYNPPD 436
||||| :||:|||||:| :| :||:| :||:| :||:| :||:| :||:|
Db 335 YKPYDAEALGLHDYCDLIKHPMDMSTTKSLKLEAREYDAQOFGADVRLMFSCYKYNPPD 394
||||| :||:|||||:| :| :||:| :||:| :||:| :||:| :||:|
QY 437 HEVYMARKLQDVEMFEMFAKMPDEPEEPVAVSSPVPPTTKVVPSSSDSSDSSDSS 496
||||| :||:|||||:| :| :||:| :||:| :||:| :||:| :||:|
Db 395 HEVYMARKLQDVEMFEMFAKMPDEPEEPVAVSSPVPPTTKVVPSSSDSSDSSDSS 448
||||| :||:|||||:| :| :||:| :||:| :||:| :||:| :||:|
QY 497 DSSTDDSEERAEQRLAEQLKAVHQLAALSOPOQONKPKK-----KEKDKKEKKEK 550
||||| :||:|||||:| :| :||:| :||:| :||:| :||:| :||:|
Db 449 DSGSSDSEERAEQRLAEQLKAVHQLAALSOQPNKPKKKEKKEKKEKKEKKEK 508
||||| :||:|||||:| :| :||:| :||:| :||:| :||:| :||:|
QY 551 HKREEVEENKSKS-----AKEPPPKPKKNNSSNSVSKPEAPMKPKPPPTYESEEDK 606
||||| :||:|||||:| :| :||:| :||:| :||:| :||:| :||:|
Db 509 EKHVKAEKKAAPPAQOAKKAPAKKANSTTTAGRLKKGKQASASYSDEEE 568
||||| :||:|||||:| :| :||:| :||:| :||:| :||:| :||:|
QY 607 CKPMSYEKQLSLDINKLPEKLGVRVHIQSRPSKNSNPDEIDFETLKPSTLRE 666
||||| :||:|||||:| :| :||:| :||:| :||:| :||:| :||:|
Db 569 GLPMSYEKQLSLDINKLPEKLGVRVHIQSRPSKNSNPDEIDFETLKPSTLRE 628
||||| :||:|||||:| :| :||:| :||:| :||:| :||:| :||:|
QY 567 L 667
: ||||| :||:|||||:| :| :||:| :||:| :||:| :||:| :||:|
Db 629 L 629

RESULT 3
BRD2_HUMAN STANDARD; PRT; 801 AA.
ID AC P25440;
DT 01-MAY-1992 (Rel. 22, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Bromodomain-containing protein 2 (RING3 protein).
GN BRD2 OR RING3 OR KIAA9001.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=T-cell;
RA Beck S., Hanson I., Kelly A., Pappin D.J.C., Trowsdale J.;
RT "A homologue of the Drosophila female sterile homeotic (fsh) gene in
RL the class III region of the human MHC.";
RN DNA Seq. 2:203-210(1992).
RN [2]
RP REVISIONS TO N-TERMINUS.
RX MEDLINE=96376536; PubMed=8781126;
RA Thorpe K.L., Abdulla S., Kaufman J., Trowsdale J., Beck S.;
RT "Phylogeny and structure of the RING3 gene.";
RN Immunogenetics 44:391-396(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RA Nomura N., Miyajima N., Sazuka T., Tanaka A., Kawarabayashi Y.,
RA Sato S., Nagase T., Seki T., Ishikawa K.-I., Tabata S.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- SIMILARITY: CONTAINS 2 BROMODOMAINS.
CC -1- SIMILARITY: CONTAINS 1 ET DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
or send an email to license@isb-sib.ch).
CC EMBL; X62083; CAA43996.1; -
CC DR EMBL; M80613; AAA68890.1; ALT_INIT.
DR EMBL; D42040; BAA07641.1; -
DR PIR; S18860; S18860.
DR PIR; S40781; S40781.
DR HSSP; Q92831; IB91.
DR Genew; HGNC:1103; BRD2.
DR MIM; 601540; -
DR InterPro; IPR001487; Bromodomain.
DR Pfam; PF00439; bromodomain; 2.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 2.
DR PROSITE; PS00633; BROMODOMAIN_1; 2.
DR PROSITE; PS00014; BROMODOMAIN_2; 2.
KW Bromodomain; Repeat; Nuclear protein.
FT DOMAIN 91 163 BROMODOMAIN 1.
FT DOMAIN 364 436 BROMODOMAIN 2.
FT DOMAIN 476 515 GLU/SER-RICH.
FT DOMAIN 544 566 ARG/LYS-RICH (HIGHLY BASIC).
FT DOMAIN 775 801 SER-RICH.
FT DOMAIN 555 559 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 638 801 ET DOMAIN.
FT DOMAIN 61 64 POLY-PRO.
FT DOMAIN 492 506 POLY-GLU.
FT DOMAIN 551 559 POLY-LYS.
FT DOMAIN 634 638 POLY-GLU.
FT DOMAIN 775 793 POLY-SER.
SQ SEQUENCE 801 AA; 88060 MW; 9A075EEB13507D8E CRC64;

Query Match 51.9%; Score 1888; DB 1; Length 801;
Best Local Similarity 56.1%; Pred. No. 5e-73;
Matches 401; Conservative 82; Mismatches 138; Indels 94; Gaps 21;

QY 5 SGPTRLRLNLPVMDGLETSMSTTQAQAQOPAPANAATNPPPTSNPNKPKQTNLQ 64
: ||||| :||:|||||:| :| :||:| :||:| :||:| :||:| :||:|
Db 27 AAPGRIRKPSLLYEGFESPTMASVPA-LQLTPA-----NPPPEVSNPKPGRTNLQ 80
: ||||| :||:|||||:| :| :||:| :||:| :||:| :||:| :||:|
QY 65 YLLRVVLTKLWKHQFANFPQOPVDVAVKLNLPDYDKIITKTPMDGTIKRLENNYVNAOE 124
||||| :||:|||||:| :| :||:| :||:| :||:| :||:| :||:|
Db 81 YLHKVVMKALWKHQFANFPQOPVDVAVKLNLPDYDKIITKTPMDGTIKRLENNYVNAOE 140
||||| :||:|||||:| :| :||:| :||:| :||:| :||:| :||:|
QY 125 CIQDFMTFTNCYIYNKPGDDIVLMALEKLFLOKINELPTEETEIMIVQAG---RG- 180
||||| :||:|||||:| :| :||:| :||:| :||:| :||:| :||:|
Db 141 CMQDFMTFTNCYIYNKPGDDIVLMALEKLFLOKINELPTEETEIMIVQAG---RG- 200
||||| :||:|||||:| :| :||:| :||:| :||:| :||:| :||:|
QY 181 -----RGRKETGTAKPGVSTVPNTQASTPQOTQTPQPNPPVQATPHFPFVTPDLIVQ 235
: ||||| :||:|||||:| :| :||:| :||:| :||:| :||:| :||:|
Db 201 KLAALQGSVTAHQVPAVSSVSHLTALYTPPPEI-----PTVLNIPHP-----SVIS 247
: ||||| :||:|||||:| :| :||:| :||:| :||:| :||:| :||:|
QY 236 TPVMTVVPVPPQLQTPFPVPPQPPQPPAPAPQPVQSHPPPIAATPQPVKTKGVRKADTT 295
: ||||| :||:|||||:| :| :||:| :||:| :||:| :||:| :||:|
Db 248 SPLKLSLH---SAGPPLAVTAAPPA-----QPLAKKKGVRKADTT 286
: ||||| :||:|||||:| :| :||:| :||:| :||:| :||:| :||:|
QY 296 TPTTIDDIHEP---PSLPP---EPKTKL-GORRESSRPVKKPKDVPDSQQHPAPEKSK 349
||||| :||:|||||:| :| :||:| :||:| :||:| :||:| :||:|
Db 287 TPTP-TAILAPGSPASPGSLPEKKAARLPPMRRESGRPIKPPKDLDPDSQQHSSKKG 345
: ||||| :||:|||||:| :| :||:| :||:| :||:| :||:| :||:|
QY 350 VSEOLKCCSGILKEMFAKHAAYAWPFKPYVDVVALGLHDYCDLIKHPMDMSTTKSLEA 409
||||| :||:|||||:| :| :||:| :||:| :||:| :||:| :||:|
Db 346 LSEOLKCCSGILKEMFAKHAAYAWPFKPYVDVVALGLHDYCDLIKHPMDMSTTKSLEA 405
||||| :||:|||||:| :| :||:| :||:| :||:| :||:| :||:|
QY 410 REYRDAQEFADVRLMFSCYKYNPPDHEVYMARKLQDVEMFEMFAKMPDEPEEPVAVS 469
||||| :||:|||||:| :| :||:| :||:| :||:| :||:| :||:|
Db 406 RYRDAQEFADVRLMFSCYKYNPPDHEVYMARKLQDVEMFEMFAKMPDEPEEPVAVS 465
||||| :||:|||||:| :| :||:| :||:| :||:| :||:| :||:|
QY 470 SPVAVPP-TKVAPSSDSSDSSDSS-----SDSDSDSDSEERAEQRLAEQ 516
: ||||| :||:|||||:| :| :||:| :||:| :||:| :||:| :||:|
Db 466 STAMPGLAKSSSSSESSSESSSESSSESSSESSSESSSESSSESSSESSSESS 525
: ||||| :||:|||||:| :| :||:| :||:| :||:| :||:| :||:|
QY 517 QLKAVHQLAALSOQONKPKKKEKKEK-----EKHKKEEVEENKKS-KAKEP-P 570
: ||||| :||:|||||:| :| :||:| :||:| :||:| :||:| :||:|
```





Db 928 RGSSKKPQVMN-----FDSEEDTAKPMSYDEKRLSLDLNKLPGDKLGRVVHIIQNR 982

Qy 641 EPSLKNSNPDEIDFETLKPSTREL 667  
||||:|||||

Db 983 EPSLRDSNPDEIDFETLKPSTREL 1009

RESULT 5

BDF1\_YEAST STANDARD; PRT; 686 AA.

AC P35817; Q06048;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE BDF1 protein.  
GN BDF1 OR YLR399C OR L8084.18.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN-S288C;  
RX MEDLINE=95116323; PubMed=7816623;  
RA Lygerou Z., Conesa C., Lesage P., Swanson R.N., Ruet A., Carlson M.,  
RA Sentenac A., Seraphin B.;  
RT "The yeast Bdf1 gene encodes a transcription factor involved in the  
RT expression of a broad class of genes including snRNAs";  
RL Nucleic Acids Res. 22:5332-5340(1994).  
[2]  
RN SEQUENCE FROM N.A.  
RA Roeder G.S., Chua P.R.;  
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.  
[3]  
RN SEQUENCE FROM N.A.  
RC STRAIN-S288c / AB972;  
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,  
RA Favello A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,  
RA Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,  
RA Johnston L., Langston Y., Latreille P., Le T., Mardis E., Menezes S.,  
RA Miller N., Nhan M., Pauley A., Peluso D., Rifken L., Riles L.,  
RA Talch A., Trevasakis E., Vignati D., Wilcox L., Wohlman P., Vaudin M.,  
RA Wilson R., Waterston R.;  
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.  
[4]  
RN SEQUENCE OF 471-686 FROM N.A.  
RX MEDLINE=93309467; PubMed=8321235;  
RA Widner W.R., Wickner R.B.;  
RT "Evidence that the SKI antiviral system of Saccharomyces cerevisiae  
RT acts by blocking expression of viral mRNA.";  
RL Mol. Cell. Biol. 13:4331-4341(1993).  
CC -1- FUNCTION: TRANSCRIPTION FACTOR INVOLVED IN THE EXPRESSION OF A  
CC BROAD CLASS OF GENES INCLUDING SNRNAs. REQUIRED FOR SPORULATION.  
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).  
CC -1- SIMILARITY: CONTAINS 2 BROMODOMAINS.  
CC -1- SIMILARITY: CONTAINS 1 ET DOMAIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Z18944; CAA79377.1; -  
CC EMBL; U18116; AAA89115.1; -  
CC EMBL; U19729; AAB82357.1; -  
CC EMBL; L13469; AAA35048.1; -  
CC PIR; A48140; A48140.  
CC PIR; A41801; S41801.  
CC TRANSFAC; T03204; -  
CC SGD; S0004391; BDF1.

DR InterPro: IPR001487; Bromodomain.  
DR Pfam: PF00439; bromodomain; 2.  
DR PRINTS: SM00503; BROMODOMAIN.  
DR SMART: SM00297; BROMO; 2.  
DR PROSITE: PS00633; BROMODOMAIN\_1; 2.  
DR PROSITE: PS00014; BROMODOMAIN\_2; 2.  
KW Transcription regulation; Nuclear protein; Bromodomain; Repeat;  
KW Sporulation.  
FT DOMAIN 165 237 BROMODOMAIN 1.  
FT DOMAIN 332 404 BROMODOMAIN 2.  
FT DOMAIN 522 686 ET DOMAIN.  
FT CONFLICT 8 8 Q -> LC (IN REF. 1).  
FT CONFLICT 93 94 GA -> R (IN REF. 2).  
FT CONFLICT 94 94 A -> P (IN REF. 1).  
FT CONFLICT 282 282 A -> P (IN REF. 1).  
FT CONFLICT 385 385 D -> E (IN REF. 1).  
FT CONFLICT 493 493 A -> R (IN REF. 4).  
SQ SEQUENCE 686 AA; 76978 MW; 8CCD52F41F91D0DA CRC64;  
  
Query Match 15.7%; Score 570.5; DB 1; Length 686;  
Best Local Similarity 25.5%; Pred. No. 1e-17;  
Matches 169; Conservative 91; Mismatches 204; Indels 199; Gaps 18;  
  
Qy 18 GDGLETSMSTTQAAQAPQ--PANAASINPPPPETSN-PNPKPQTQNLQYLLRVVLKTL 74  
Db 107 GQGTQKQEDLDSKQELPMVEKPEPAPPEDPMNNLPQNPPIPKHQKHALL--AIKAV 164  
Qy 75 WKHGFAMPFPQPDVAVKLNLPDYKIIKTPMDMGTIKKRLNENNYWNAQECIQDFTMET 134  
Db 165 KRLKDARFELQPDVVKLDIPFFENYIKRPMDLSTIERKLVNGAYEVEPQITEDFNLMVN 224  
Qy 135 NCYIYNKPGDDIVLMAEALKFLQKINELPTEETETIMVOAKGRGRKETGTAKPGVS 194  
Db 225 NSIKFNGPNAGISQAMARNIOASFEEKMLNMPAKDA----- 259  
Qy 195 TVPNTTQASTPPQQTQTPQNPVQATPHPPVATPDLIQTVPMVTVVPPQPLTPPPVP 254  
Db 260 ----- 259  
Qy 255 PQQPPAPAPQPVQSHPPPIIAATPQVTKKGVKKADTTPTTIDPIHEPPSLPPEPK 314  
Db 260 -----PPVIA-----KGRSSAQEDAPIVIRRAQTHNGRP----- 289  
Qy 315 TTLGQRRESSRPVKKPK-KDV-PDSQOHPAPEKSVSEOLKCCSGILKEMFAKHAAY 372  
Db 290 -----KRTTHPPKSDIYPYESSKP---KSKRLQQAAMKFCQSVLKELMAKHASY 336  
Qy 373 AWPEYKPDVVENLGLHDYDCIHKHPMDMSTIKSLEAREYDAQEFQADYVLMESNCYKY 432  
Db 337 NYFLEPVDVPMNLPITYFDYVKEPMDLGTAKKLDNDWQYQTMEDFEDRVLPVFKNCYTF 396  
Qy 433 NPDHEVYVAMARKLDQYFEMRFAPKMPDEPEPVAVVSPVPPPTKVAVPSSSDSSDS 492  
Db 397 NPDGTIVMGMHRLLEEVFNKWDNR-----PNDYDSDS 431  
Qy 493 SSDSDSDTDDSEERAQRLAELOE-----OLKAVHQLAALS-QPOONPKPKKEDKKE 545  
Db 432 DSRTQGDYDDYSEYSE--SDIDETIITNPAIOVLEBQLARKMKVELQQLKKQLEKIRKE 489  
Qy 546 KKEKHKKRKEEVEENKSKAKEPPPKTKKNNSNSNVSKKEPAPMKSKPPPYESEED 605  
Db 490 RLARGSKK-----RGRKSKG-----RSGKNASSKGRDRKK-----N 522  
Qy 606 KCKP-MSVEEKROLSLDINKLPGKLGKRVVHIQTOSREPSLKNNSNPDEIDFETLKPSTL 664  
Db 523 KLKTVVYDMKRIITERINDLPSTSKLERAIDIIKSKPNIT--SEDDVEVLDTLDNHTI 580  
Qy 665 REL 667  
Db 581 LTL 583  
  
RESULT 6

CBP\_MOUSE  
ID CBP\_MOUSE STANDARD; PRT: 2441 AA.  
AC P45481;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE CREB-binding protein.  
GN CREBBP OR CBP.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=94019866; PubMed=8413673;  
RA Chivria J.C., Kwok R.P.S., Lamb N., Hagiwara M., Montminy M.R.,  
RA Goodman R.H.;  
RT \*Phosphorylated CREB binds specifically to the nuclear protein CBP.\*;  
RL Nature 363:855-859(1993).  
CC -!- FUNCTION: MEDIATES CAMP-GENE REGULATION BY BINDING SPECIFICALLY TO  
CC PHOSPHORYLATED CREB PROTEIN. ACTING AS A COACTIVATOR, CBP AUGMENTS  
CC THE ACTIVITY OF PHOSPHORYLATED CREB TO ACTIVATE TRANSCRIPTION OF  
CC CAMP-RESPONSIVE GENES.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- SIMILARITY: CONTAINS 1 BROMODOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 Z2-TYPE ZINC FINGER.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: S66385; AAB28651.1; -  
DR TRANSFAC; T01318; -  
DR MGD; MGI:1098280; Crebbp.  
DR InterPro; IPR001487; Bromodomain.  
DR InterPro; IPR003101; KIX.  
DR InterPro; IPR000197; TAZ\_finger.  
DR InterPro; IPR000433; Znf\_Z2.  
DR Pfam; PF00439; bromodomain; 1.  
DR Pfam; PF00569; ZF\_1.  
DR Pfam; PF02135; ZF-TAZ; 2.  
DR Pfam; PF02172; KIX; 1.  
DR PRINTS; PR00503; BROMODOMAIN.  
DR SMART; SM00297; BROMO; 1.  
DR SMART; SM00291; Znf\_Z2; 1.  
DR PROSITE; PS00633; BROMODOMAIN\_1; 1.  
DR PROSITE; PS50014; BROMODOMAIN\_2; 1.  
DR PROSITE; PS01357; ZF\_Z2\_1; 1.  
DR PROSITE; PS0135; ZF\_Z2\_2; 1.  
KW Transcription regulation; Nuclear protein; Activator; Bromodomain;  
KW Zinc-finger.  
FT DOMAIN 1104 1176 BROMODOMAIN.  
FT ZN\_FING 1702 1745 Z2-TYPE.  
FT DOMAIN 1062 1065 POLY-GLU.  
FT DOMAIN 1556 1563 POLY-GLU.  
FT DOMAIN 1944 1949 POLY-PRO.  
FT DOMAIN 1968 1971 POLY-GLN.  
FT DOMAIN 2082 2086 POLY-GLN.  
FT DOMAIN 2200 2216 POLY-GLN.  
FT DOMAIN 2296 2299 POLY-GLN.  
SQ SEQUENCE 2441 AA; 265474 MW; 0AB8028C3112F419 CRC64;  
Query Match 8.9%; Score 323.5; DB 1; Length 2441;  
Best Local Similarity 23.4%; Pred. No. 7.9e-07;  
Matches 135; Conservative 72; Mismatches 177; Indels 193; Gaps 27;  
QY 9 TRLRLNLPVYMGDGLFTSQMSTTQAQAQPOPANAASNPPT--PETSNNPKPKRQTNLQYL 66  
||| ::| :||| ||| ||| | :|

Db 670 TRLHKQGIIGN-----QPALPASGAQPPVIPPAAQSVRRPN----- 704  
QY 67 LRVVLKTLWKHQFAMFPQPPQVDAVKLN--LPDYKKIITPMDMGTIK-----KR 113  
Db 705 -----GPLDLPVNRMQVSGMNSF-----NPMSLGNVOLPOAPMGPRASP 745  
QY 114 LENNYWNAQECI-----QDFNTMTFCYIYNKPGDDIVLMAEA-LEKILFLQKI 161  
Db 746 MNHVSOMNSMASVPGMAISPSRMPQPPNMGTHA---NN-----IMAAQPTNQFLPQ- 795  
QY 162 NELPTEETIIMIVQAKGRGRKKTGTA---KPGVS-----TVPNTTOASTPPQTQTP 211  
Db 796 NOFPS--SSGAMSVNSVGMQPAAGVSGQGEPPGAALPNLMLAPQASQLPCPPVTQSP 854  
QY 212 -QPNPPP-----VOATPHP-PPAVTLPDILVQTPVMTVVPPQPL---OTPPPPV- 254  
Db 855 LHPTPPASTAAGMPSLOHPTAGMTP---PQPAAPTQPTPTPVSSGGTPTTTPGCVPSA 910  
QY 255 -----PQPPPPAP-----APQPVQSHPPPII----- 275  
Db 911 AQTQSTPTVQAAAAQAVTPQQTVPQPPSVATPQSSQQOQTPVHTQPPGCTPLSQAASID 970  
QY 276 --AATPQPVKTKKGVRKADITTP-----TTIDPIHEPSPSLPEPKTTKLCORRESS 325  
Db 971 NRVPTSTVTSAETSSQQPGDPVPMLEKTEVQTDADPEPTESKGEPRSEMMEEDLOGS 1030  
QY 326 RPVK-----PPKQDVPDSQQHPAPE-----KSKVSEQLKCCGILKEMFAKK- 368  
Db 1031 SQVKEETDTTEKSEPMEEKKEPVKEAKEEENSNDTASQTSQSPQPKKIFKPEE 1090  
QY 369 -----HAAY-----AWPFYKPDVVEALGLHDYCDIHKPDMSTKSKLEAREYD 414  
Db 1091 LRQALMPTLEALYRDPESLPPRPQVDPQLLIGIPDYFDIVKNPMDLSTIKRKLDTGQYOE 1150  
QY 415 AQEGADVRLMFSNCKYKNPPDPHEVVMARKLDQYFE 451  
Db 1151 PQWYVDDVRLMFWNNAWLNRYKTSRVYKFCSKLAEFVE 1187  
RESULT 7  
CBP\_HUMAN  
ID CBP\_HUMAN STANDARD; PRT: 2442 AA.  
AC Q92793; Q16376; O00147;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE CREB-binding protein.  
GN CREBBP OR CBP.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97385172; PubMed=9238046;  
RA Sobulo O.M., Borrow J., Tomek R., Reshmi S., Harden A.,  
RA Schlegelberger B., Housman D., Doggett N.A., Rowley J.D.,  
RA Zelezniak-Le N.J.;  
RT \*MLL is fused to CBP, a histone acetyltransferase, in therapy-related  
RT acute myeloid leukemia with a t(11;16)(q23;p13.3).\*;  
RL Proc. Natl. Acad. Sci. U.S.A. 94:8732-8737(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97321049; PubMed=9177780;  
RA Giles R.H., Petrij F., Dauwerse H.G., den Hollander A.I.,  
RA Lushnikova T., van Ommen G.J.B., Goodman R.H., Deaven L.L.,  
RA Doggett N.A., Peters D.J.M., Breuning M.H.;  
RT \*Construction of a 1.2-Mb contig surrounding, and molecular analysis  
RT of, the human CREB-binding protein (CBP/CREBBP) gene on chromosome  
RL 16p13.3\*;  
RN Genomics 42:96-144(1997).  
RP [3]  
RP SEQUENCE OF 1-405 FROM N.A.

RA MEDLINE=96376968; PubMed=8782817;  
RA Borrow J., Stanton V.P., Andresen J.M., Becher R., Behm F.G.,  
RA Chaganti R.S.K., Clavin C.I., Distche C., Dube I., Frischauf A.M.,  
RA Horman D., Micelman F., Volinia S., Watmore A.E., Housman D.E.;  
RT "The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses  
RT a putative acetyltransferase to the CREB-binding protein. #;  
RL Nat. Genet. 14:33-41(1996).  
RN [4]  
RN INTERACTION WITH PCAF.  
RX MEDLINE=96300317; PubMed=8684459;  
RX Yang X.-J., Ogrzyzko V.V., Nishikawa J.-I., Howard B.H., Nakatani Y.;  
RA "A p300/CBP-associated factor that competes with the adenoviral  
RT oncoprotein ELA".  
RL Nature 382:319-324(1996).  
CC -|- FUNCTION: MEDIATES CAMP-GENE REGULATION BY BINDING SPECIFICALLY TO  
CC PHOSPHORYLATED CREB PROTEIN. ACTING AS A COACTIVATOR, CBP AUGMENTS  
CC THE ACTIVITY OF PHOSPHORYLATED CREB TO ACTIVATE TRANSCRIPTION OF  
CC CAMP-RESPONSIVE GENES.  
CC -|- SUBUNIT: INTERACTS WITH SMAD1, SMAD2, SMAD3 AND PCAF.  
CC -|- SUBCELLULAR LOCATION: Nuclear.  
CC -|- DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY CHROMOSOMAL TRANSLOCATIONS  
CC T(8;16)(P11;P13) INVOLVING CBP AND MOZ, AND T(11;16)(Q23;P13.3)  
CC INVOLVING CBP AND MLL.  
CC -|- DISEASE: DEFECTS IN CREBBP ARE THE CAUSE OF RUBINSTEIN-TAYBI  
CC SYNDROME (RITS), A DISORDER CHARACTERIZED BY CRANIOFACIAL  
CC ABNORMALITIES, BROAD THUMBES, BROAD BIG TOES, MENTAL RETARDATION  
CC AND A PROPENSITY FOR DEVELOPMENT OF MALIGNANCIES.  
CC -|- SIMILARITY: CONTAINS 1 BROMODOMAIN.  
CC -|- SIMILARITY: CONTAINS 1 Z2-TYPE ZINC FINGER.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
DR EMBL: U47741; AAC51770.1; -;  
DR EMBL: U85962; AAC51331.1; -;  
DR EMBL: U89354; AAC51339.1; -;  
DR EMBL: U89355; AAC51340.1; -;  
DR TRANSFAC: T02214; -;  
DR Genew: HGNC:2348; CREBBP.  
DR MIM: 600140; -;  
DR InterPro: IPR001487; Bromodomain.  
DR InterPro: IPR003101; KIX.  
DR InterPro: IPR000197; TAZ\_finger.  
DR InterPro: IPR000433; Znf\_Z2.  
DR Pfam: PF00439; bromodomain; 1.  
DR Pfam: PF00569; Z2; 1.  
DR Pfam: PF02135; zf-TAZ; 2.  
DR Pfam: PF02172; KIX; 1.  
DR PRINTS: PR00503; BROMODOMAIN.  
DR SMART: SM00297; BROMO; 1.  
DR SMART: SM00291; Znf\_Z2; 1.  
DR PROSITE: PS00633; BROMODOMAIN\_1; 1.  
DR PROSITE: PS00014; BROMODOMAIN\_2; 1.  
DR PROSITE: PS01357; ZF\_Z2\_1; 1.  
DR PROSITE: PS01355; ZF\_Z2\_2; 1.  
KW Transcription regulation; Nuclear protein; Activator; Bromodomain;  
KW Chromosomal translocation; Zinc-finger.  
FT ZN\_FING 1701 1744  
FT DOMAIN 363 430  
FT DOMAIN 452 683  
FT DOMAIN 1103 1175  
FT DOMAIN 1061 1064  
FT DOMAIN 1199 1487  
FT DOMAIN 1555 1562  
FT DOMAIN 1675 1849  
FT DOMAIN 1943 1948  
FT DOMAIN 1967 1970  
FT DOMAIN 2081 2085

FT DOMAIN 2199 2216  
FT DOMAIN 2245 2248  
FT DOMAIN 2297 2300  
FT CONFLICT 1513 1725  
FT CONFLICT 1724 1725  
FT CONFLICT 1770 1770  
FT CONFLICT 1789 1789  
FT CONFLICT 1812 1812  
SQ SEQUENCE 2442 AA; 265336 MW; 42D084619475F3D2 CRC64;  
Query Match 8.5%; Score 308; DB 1; Length 2442;  
Best Local Similarity 26.1%; Pred. No. 3.6e-06;  
Matches 109; Conservative 49; Mismatches 132; Indels 128; Gaps 19;  
QY 148 LMAEA-LEKIFLQKINELPTEETIMIVQAKGRGRKGTAK---PGVSTV----- 196  
DB 783 MMAQAPAAQSOLFQ-NQFPSSSGAMSV-----GMCQPPAQGTGVSQGVGAALPNPLNMLG 837  
QY 197 PNTQASTPPQTQTP-QPNPPVQATPHPPFVNTPDILIVTPVMTVVPQP-LQTPPVP 254  
DB 838 PQASQLPCPVTOSPLHPTPPASTA-----AGMPSLOHTTP-PGMTTPQPPAAPTQSTP 891  
QY 255 -----POQPPAPAPQPVQSHPIIAA-----TPQ---PVKTKGVRKADTTTPTT 299  
DB 892 VSSGGQTPTTPGVSPTSATQSTPTVQAAQAQVTPQTPVQPPSVATPQSSQQQPT- 950  
QY 300 IDPIH-EPPSLPPEPKTKLGQRRESSRPV-----KPKKQDVP-----DSQQ 340  
DB 951 --PVHAPPCTPLSQAASIDNRVPTPSSVASAETNSQPCDPVPLVEMKTETQAEDEP 1008  
QY 341 HPAEK-----SSKVSQELKCSGILKE----- 363  
DB 1009 DPGESKGEPRSEMEEDLQASQVKEETDAEQSEPMYDEKKPEYKVEVEEESSEN 1068  
QY 364 -----MFAKHAAY-----AWPFYKVDVDEALGLHDYCDI 393  
DB 1069 GTASQSTSPSQPRKKTFKPEELQALMPTLEALYRQDPESLPRQPVDPQLGLPDTFDI 1128  
QY 394 IKHPMDMTSKLEAREYRDAQEFQADVRLMFSNCYKYNPPDPHEVVAMARKLDQVFE 451  
DB 1129 VKNPMDLSTIKRLDGTGOYQEPQYVDDVWLMFNNAWLYNRKTSRYVKFCSKLAEVFE 1186  
RESULT 8  
ID P300\_HUMAN STANDARD; PRT; 2414 AA.  
AC Q09472;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE ELA-associated protein p300.  
GN EP300 OR P300.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
SEQUENCE FROM N.A.  
RP MEDLINE=95011587; PubMed=7523245;  
RX Eckner R., Ewen M.E., Newsome D., Gerdes M., Decaprio J.A.,  
RA Lawrence J.B., Livingston D.M.;  
RT "Molecular cloning and functional analysis of the adenovirus ELA-  
RT associated 300-kD protein (p300) reveals a protein with properties of  
RT a transcriptional adaptor".  
RL Genes Dev. 8:869-884(1994).  
RN [2]  
INTERACTION WITH PCAF.  
RX MEDLINE=96300317; PubMed=8684459;  
RA Yang X.-J., Ogrzyzko V.V., Nishikawa J.-I., Howard B.H., Nakatani Y.;  
RT "A p300/CBP-associated factor that competes with the adenoviral  
RT oncoprotein ELA".  
RL Nature 382:319-324(1996).  
CC -|- FUNCTION: Probable transcriptional adapter required for the

```
CC activity of certain complex transcriptional regulatory elements.  
CC May have a function in cell cycle regulation. Binds to and may be  
CC involved in the transforming capacity of the adenovirus EIA  
CC protein.  
CC -|- SUBUNIT; Interacts with PCAF.  
CC -|- SUBCELLULAR LOCATION; Nuclear.  
CC -|- SIMILARITY; CONTAINS 1 BROMODOMAIN.  
CC -|- SIMILARITY; CONTAINS 1 ZF-TYPE ZINC FINGER.
```

---

```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).
```

---

```
DR EMBL; U01877; AAA18639.1; --  
DR TRANSFAC; T01427; --  
DR Genew; HGNC:3373; EP300.  
DR MIM; 602700; --  
DR InterPro; IPR001487; Bromodomain.  
DR InterPro; IPR003101; KIX.  
DR InterPro; IPR000197; TAZ_finger.  
DR InterPro; IPR000433; ZnF_ZZ.  
DR Pfam; PF004439; bromodomain_1.  
DR Pfam; PFO0569; zz; 1.  
DR Pfam; PF02135; zf-TAZ; 2.  
DR Pfam; PF02172; KIX; 1.  
DR PRINTS; PR00503; BROMODOMAIN.  
DR SMART; SM00297; BROMO; 1.  
DR SMART; SM00291; ZnF_ZZ; 1.  
DR PROSITE; PS00633; BROMODOMAIN_1; 1.  
DR PROSITE; PS00014; BROMODOMAIN_2; 1.  
DR PROSITE; PS01357; ZF_ZZ_1; 1.  
DR PROSITE; PS0135; ZF_ZZ_2; 1.  
KW Transcription regulation; Nucleolar protein; Bromodomain; Cell cycle;  
KW Zinc-finger.
```

---

```
KW DOMAIN      11          17    NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT FT         1067       1139   BROMODOMAIN.  
FT FT         1572       1818   BINDING REGION FOR E1A ADENOVIRUS.  
FT FT         1664       1707   ZZ-TYPE.  
FT FT           797        800   POLY-SER.  
FT FT         1519       1526   POLY-GLU.  
FT FT         2066       2069   POLY-GLN.  
FT FT         2190       2195   POLY-LIN.  
SQ SEQUENCE     2414 AA; 264143 MW; 68FF909EE4B9D693 CRC64;
```

---

```
Query Match            8.4%; Score 305; DB 1; Length 2414;  
Best Local Similarity 22.9%; Pred.No.4.7e-06;  
Matches 133; Conservative 65; Mismatches 165; Indels 218; Gaps 24;  
QY 9 TRIG-RNLPMVGDCLETSOMSTQAQAPQPANAASTNNPPPETSS-----NPNRPK-- 57  
DB ||| :||: |:::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||  
650 TLKQKNPLNAAGMVPVSMNGPNMGQQP--GMTSNGLPLPDPSMTIRSGVSPNQMMPRIT 707
```

---

```
QY 58 -----ROYNQLLYLRVVLTWKHQFAWPFPQPVDAVKLNLPDY 97  
DB ||| ||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||  
708 PQSLNQFGMSMAQPIIVRQRPFLQH-----HCQLAQP----- 742
```

---

```
QY 98 YKIKTMDMGTIKKRLANNYYNAQCIEDFTNTFCNYIYNKPGGDDIVLMAEALEKL 157  
DB || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||  
743 -GALNPNMGYPGRWOOPNSQGELPQ-----TFPS----- 772
```

---

```
QY 158 LQKINELPTETEIMIVQAKRG---GRKRKETAKPGVTVDNTQTASTPDTTPQPOP- 213  
DB || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||  
773 -QGMN----VTNIPLAPSQSQAFAVSQAQSSSCPCVNPSIMPGPSQGSIIHCPQLPQPA 826
```

---

```
QY 214 ---NPPPVOA---THPHPEAVTVLDLVDPVMVTVPPOPLQTPPPVPVPOQ----- 258  
DB || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||  
827 LHONSFSVPESRTPTPHHTP---PSGAQQOAPPATTI-PADVPITPPAMPGPQSOALHPHP 882
```

---

```
QY 259 ----PPAPAPAPOGVSHSPIIAATQPVRTKKGVRKA DTPTTTIDPLEHEPSLPPEP 313
```

colon, spleen, kidney.

-1- MISCELLANEOUS: Candidate gene for the Cat Eye Syndrome (CES), a developmental disorder associated with the duplication of a 2 Mb region of 22q11.2. Duplication usually takes in the form of a supernumerary bisatellited isodicentric chromosome, resulting in four copies of the region (represents an inv dup(22)(q11)). CES is characterized clinically by the combination of coloboma of the iris and anal atresia with fistula, downslanting palpebral fissures, preauricular tags and/or pits, frequent occurrence of heart and renal malformations, and normal or near-normal mental development.

-1- SIMILARITY: CONTAINS 1 BROMODOMAIN.

-----

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; AF336133; AAK15343.1; -  
 EMBL; AB051527; BAB21831.1; -  
 EMBL; AF411609; AAL07393.1; -  
 HSSP; Q92831; 1B91  
 Genew; HGNC:1840; CECR2.  
 InterPro; IPR001487; Bromodomain.  
 Pfam; PF00439; bromodomain; 1.  
 PRINTS; PR00503; BROMODOMAIN.  
 SMART; SM00297; BROMO; 1.  
 PROSITE; PS00633; BROMODOMAIN\_1; 1.  
 PROSITE; PSS0014; BROMODOMAIN\_2; 1.  
 KW Alternative splicing.  
 FT DOMAIN 451 521 BROMODOMAIN.  
 FT DOMAIN 333 337 POLY-GLU.  
 FT DOMAIN 611 614 POLY-SER.  
 FT DOMAIN 1250 1253 POLY-PRO.  
 FT VARSPIC 291 318 MISSING (IN ISOFORM B).  
 FT VARSPIC 519 526 EYKMSDN -> GKGRSLC (IN ISOFORM B).  
 FT VARSPIC 527 1484 MISSING (IN ISOFORM B).  
 FT CONFLICT 370 389 MISSING (IN REF. 2).  
 FT CONFLICT 1029 1029 C -> S (IN REF. 2).  
 FT CONFLICT 1045 1045 R -> W (IN REF. 2).  
 SQ SEQUENCE 1484 AA; 164214 MW; 049AA844E51AF63F CRC64;

Query Match 8.3%; Score 300.5; DB 1; Length 1484;  
 Best Local Similarity 20.9%; Pred. No. 4.6e-06;  
 Matches 140; Conservative 85; Mismatches 265; Indels 179; Gaps 26;

QY 47 PETS--NPKNKRTQWLOQLLRV-----VLTKLKHQFAPFPQDPVDAVKLNLP 95  
 DB 412 PPELSHLDNPSMKREKTKDLFLDDFTAMVKVLDVAVKADSWPFLFPV--ESYAP 469  
 QY 96 DYKIIITPMDMGTIKKRLNNYVNAQECIQDFTMTNCTIYNKPGDDIVLMAEALEK 155  
 DB 470 NYQIIKAPMDISSMEKGLNGLYCTKEEFVNDMKTFRNCRKINGESSEYTKMSDLR 529  
 QY 156 LFQK-KINELPTE---ETEIMIVQAKGRGRKRTGTA----- 189  
 DB 530 CFHAMKHPFGEDGDTDEEFWIREDEKREKRSRACRSGSHVWTRSDRPEGSSRRKQP 589  
 QY 190 -KPGVSTVNTQASTPQTPQPPPPVQ-----ATPHPPP-AVTPDLIVQTPVMT 240  
 DB 590 MNGGKSLPPTTRAPSSGDDQSSSTQPPREVGTSGRGSFSLHLCGGTGS--QAPFLN 646  
 QY 241 VVPQPLQTPPPVP---POQPPPPA-----PAP-OPVQSHPPILAAATPQPVKTK 285  
 DB 647 -----QMPAPVGTGRLGSDPAFLYGSSGVLEPHGPEVQQRQPTWQPPVGNLSL 699  
 QY 286 KGVKKADTTTTPTTIDPIHEPPSLPPPKTKLQ-----RRE 323  
 DB 700 RG-PRLTPEEKQMCGLTHLSNMGPHPGSLQLQISGPSQDSMYAPAQFQPGFIPRH 758

QY 324 SSRPVKPKKVDPSQQHQAPEKSKVSEQLKCCSGILKEMFAKKAAY----- 372  
 DB 759 GGAPARPP--DFPESSEIP-----PSHMYRSYKILNRVHSAVNGNHGATNGPLGPDEKP 812  
 QY 373 ---AWPFKYPVDVEALGLHDYCDIHKHPMDMSTKSKLEAREYRDAQEF---GADVRLM 425  
 DB 813 HLGPGSHOP---RTLG-----HYMDSRVMRPPVPPNQWTEOSGFLPHGVPSSGYM 860  
 QY 426 FSNC-----YKNNP-----DHEVVAMARKLDQVFEFR 454  
 DB 861 RPPCKSAGHRLQPPVPAPSSLFGAPALRGVQGGSDMSDPSPEMIAMQQLSSRVCPGCV 920  
 QY 455 AKMDEPEEPVAVSSPAVPPTKVAVPSSSDSSSDSSSDSSDSEERAQRLAEL 514  
 DB 921 PYHPQAPHLPLPGFPQVAHPMSVTVSAPKALGNPCRPENSEAQEPENDQAPLPLGL 980  
 QY 515 QEQLKAVHEQ---LAALSQP-----QONKPKKKEKDKKKKKKKHKKKEVEENK 561  
 DB 981 EEPGPGVGTSGVYLTQLPHPTPLQTDCTROSSPOERTVGPPELKS---SCSESADNCK 1037  
 QY 562 KSKAKEPPPP 570  
 DB 1038 AMKGKNRP 1046

RESULT 10  
 NFHL\_HUMAN  
 ID NFHL\_HUMAN STANDARD; PRT; 1020 AA.  
 AC P12036;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Neurofilament triplet H protein (200 kDa neurofilament protein)  
 DE (Neurofilament heavy polypeptide) (NF-H).  
 GN NEFH OR NEFH.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88328981; PubMed=3138108;  
 RA Lees J.F., Shneidman P.S., Skuntz S.F., Carden M.J., Lazzarini R.A.;  
 RT "The structure and organization of the human heavy neurofilament  
 subunit (NF-H) and the gene encoding it.";  
 RL EMBO J. 7:1947-1955(1988).  
 CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,  
 AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.  
 CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT  
 CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.  
 CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P. NEF IS  
 CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS  
 CC THOUGHT THAT PHOSPHORYLATION OF NEF RESULTS IN THE FORMATION OF  
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE  
 CC OF AXONAL CALIBER.  
 CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING  
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H). THE  
 CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND  
 CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.  
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; X15306; CAA33366.1;  
 EMBL; X15307; CAA33366.1; JOINED.  
 EMBL; X15308; CAA33366.1; JOINED.  
 EMBL; X15309; CAA33366.1; JOINED.

DR PIR: S00979; QFHUH.  
DR Genew; HGNC:7737; NEFH.  
DR MIM; 162230; -.  
DR InterPro; IPR001664; IF.  
DR Pfam; PF00038; filament; 1.  
DR PROSITE; PS00226; IF; 1.  
KW Intermediate filament; Coiled coil; Neurone; Phosphorylation.  
FT DOMAIN 1 100 HEAD.  
FT DOMAIN 101 413 ROD.  
FT DOMAIN 414 1020 TAIL.  
FT DOMAIN 101 132 COIL 1A.  
FT DOMAIN 133 145 LINKER 1.  
FT DOMAIN 146 244 COIL 1B.  
FT DOMAIN 245 266 LINKER 12.  
FT DOMAIN 267 288 COIL 2A.  
FT DOMAIN 289 292 LINKER 2.  
FT DOMAIN 293 413 COIL 2B.  
SQ SEQUENCE 1020 AA; 111780 MW; 1177C9DCB3DCF1D4 CRC64;  
  
Query Match 7.4%; Score 267.5; DB 1; Length 1020;  
Best Local Similarity 21.0%; Pred. No. 8e-05;  
Matches 154; Conservative 118; Mismatches 316; Indels 145; Gaps 29;  
  
QY 19 DGLSTQKSTTQAQAPQANAASTNPPPTSNPKPQTNQLOQLLRVVLKTLWKHQ 78  
DB 309 DAMSAQEITEYRQLQ-----ARTLEALKSTKDSLERQSELED-----RHQ 354  
QY 79 -----FAMFPQVDAVKLN-----LPDYKIKITPM-----DMGTIKKRLNNYWN 121  
DB 355 ADIASYQEAIQ-LDALENTKWEAAQLREYQDLLNVKMAIDIEIAYRKLE-----407  
QY 122 AQECIQDNFM-FT-----NCYIYNKPGDDIVLMAEALFKLFQ 159  
DB 408 GEGRIGFGPIPLPEGLPKIPSVSTHIKVKSEKIKVKESEKTVIVEOTE--T 464  
QY 160 KINELPTEETIMIVQAGRGRETGTAKGVSTVNTTQASTPPTQTPQ-----NP 215  
DB 465 QVTEETEEKEEKEEGEEEGEEAEGG-----EETKSPPAEEAASPEKEAKS 518  
QY 216 PPVQATPHFPFPAVTPDLIVQTVMTVWPVPPQLQTPPPPPQPPAPQVQSHPII 275  
DB 519 PVKEEAKSPAEEKSPEKEAKSPAEEKSPEKAKSPEAKSPEAKSPEAKSPA 577  
QY 276 AATPOPKVT--KGVKRRADTTPTTI-DPIHEPSLPEPKTKLGGRRSSRPVKPK 332  
DB 578 VKSPEKASPAKEEAKSPAEEKSPEKASPAKEEAKSPAEEKSPEKASPAEK 637  
QY 333 KDVPDSQHPAPEKSKYSEQLKCCGILKEMFAKHAAYAWPFYKPYDVEALGLHDYCD 392  
DB 638 AKSPTEEAKSPEKASPE-----KEEAKSPEKASPVKAEA-----674  
QY 393 ITHPMDMSTIKSKLEAREYRDAQPFQADVRLMFNSCYKYNPPDHDEVVAMARKLDQVPEM 452  
DB 675 ---KSP---EAKSPVKA-EAKSPEKASPVKEEAKSPEKASPVKEEAKSPEKASPVKE 728  
QY 453 RFAKMPDPEEPV-----VAVSSPVPVPPYKVVAVPSSSDSSSDSSSDSSDSSDSSDSEE 505  
DB 729 E-AKTPEKASPVKEEAKSPEKASPEKAKTLDVKSPEAKTPEAKPEAKSPADKF---PEK 784  
QY 506 ERAQRLAELOELKAVHEQLAALSQPNQKPKKKEK---KEKKKKEKHKEEVEENKK 562  
DB 785 AKSPVKEEAKSPEKASPLKADAKAPEIKPKKEEVKSPVKEEAKPVKEEAKPVKEEAKPE 844  
QY 563 SKAPPPPKTKTKNNSSNSVSKKPPAPMKSPPTTYSSEE---DKCKPMSEYKPKRLS 619  
DB 845 EKAPATPKTEEKD-----SKKEAPKKEAPKPVKEEAKPEKPEKPEKPEKPEKPEKPE 897  
QY 620 LDINKL--PGEKGRVHIIQRESPLK-----NSNPDEIEIDFTLKPSTRELKALCHL 673  
DB 898 EDKKVKVTPPEKAPAVEKEDAKPEKTEKPEKPEKPEKPEKPEKPEKPEKPEKPEKPE 950  
QY 674 LFAEEKETFKLRK 686

Db 951 --APEKDKTKEEK 961

RESULT 11  
CBPL\_CAEEL STANDARD; PRT: 2056 AA.  
AC P34545;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Protein cbp-1.  
GN CBP-1 OR R10E11.1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M., Coulson A.,  
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fraser A.,  
RA Craxton M., Dear S., Du Z., Durbin R., Pavello A., Fraser A.,  
RA Fultton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,  
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,  
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,  
RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,  
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,  
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,  
RA Wohlman P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RL elegans.";  
RL Nature 368:32-38(1994).  
RN [2]  
RP REVISIONS, AND ALTERNATIVE SPLICING.  
RA Durbin R.;  
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- ALTERNATIVE PRODUCTS: 2 isoforms; a and b (shown here); may be  
CC produced by alternative splicing.  
CC -!- SIMILARITY: CONTAINS 1 BROMODOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 ZF-TYPE ZINC FINGER.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Z29095; CAA82353.2; -.  
CC EMBL; Z29095; CAD18875.1; -.  
CC PIR; S40713; S40713.  
CC WormPep; R10E11.1a; CE28069.  
CC WormPep; R10E11.1b; CE21117.  
CC InterPro; IPR001487; Bromodomain.  
CC InterPro; IPR003101; KIX.  
CC InterPro; IPR001230; Prenyl site.  
CC InterPro; IPR000197; Taz\_finger.  
CC InterPro; IPR000433; Znf\_ZZ.  
CC Pfam; PF00439; bromodomain; 1.  
CC Pfam; PF00569; ZF; 1.  
CC Pfam; PF02135; ZF-TAZ; 2.  
CC Pfam; PF02172; KIX; 1.  
CC PRINTS; PR00503; BROMODOMAIN.  
CC SMART; SM00297; BROMO; 1.  
CC SMART; SM00291; Znf\_ZZ; 1.  
CC PROSITE; PS00633; BROMODOMAIN\_1; 1.  
CC PROSITE; PS50014; BROMODOMAIN\_2; 1.  
CC PROSITE; PS01359; ZF\_PHD\_1; 1.  
CC PROSITE; PS01357; ZF\_ZZ\_1; 1.  
CC PROSITE; PS50135; ZF\_ZZ\_2; 1.

KW Bromodomain; Metal-binding; Zinc; Zinc-finger; Alternative splicing.  
FT DOMAIN 881 953 BROMODOMAIN.  
FT ZN\_FING 1493 1534 Z2-TYPE.  
FT DOMAIN 1687 2008 GLY/GLN-RICH.  
FT VARSPLIC 467 478 SDTQTKKCSV -> F (IN ISOFORM A).  
SQ SEQUENCE 2056 AA; 227179 MW; 949F4608C634F01 CRC64;  
  
Query Match 7.3%; Score 265; DB 1; Length 2056;  
Best Local Similarity 22.5%; Pred. No. 0.00019;  
Matches 113; Conservative 69; Mismatches 195; Indels 126; Gaps 16;  
  
QY 2 SAGSGPTRLNLPVMDGLETS-----QMSSTQAAQAPQAPANAASNPDPET 50  
DB 533 SAE-GDGLHFGSPAMTGNITSLPEGFNGNPFQNGPRPGCGNGEINLPDPDM 591  
QY 51 SNPNKPKROTNOYLRL-VYLTLMKHQPFPQFQVDAVKLNLPDYKIIKTPMDMGT 109  
DB 592 --PDCTKEWHQVTKDLRNHLVGLKVAIFPEPQEAAMNDRLK-----DLIA 637  
QY 110 IKRLENNYWNQAECLQDFTWFTNCY-----IYNKPGDDIVLMAPALEKLFLOKINEL 164  
DB 638 YARKVEKEMPESANDREYYHLAEKYYIKQLQEKKNRLNQAAGAAHQYAIPPSNEL 697  
QY 165 PTEETEIMVQAKGRGRKETCTAKPGVSTVP-----NTTQASTPPQTPTQ 212  
DB 698 -----AOMLGVEGGRSDVISESSMAVAPSOQNPMGAGPNSNMHQIIPNGOVQP 748  
QY 213 PN-----PPVQATPHFFAFTVLDIVQVMTVVPVPPQLPTPPVPPQPPAPAPQP 267  
DB 749 VNNSSTFPSSGNETPNIGASSTVSAMQ-----PKTEPMDQNTDS 789  
QY 268 VQSHPII-----AATQPVTKKGVKKKADTTTPTDTHPEPSLPPEPKTKLG- 319  
DB 790 LSSRPPTAIFGSGSSSTFAP1--MNGIVKKEE-----DPESSNQAPPSVKDKDGV 840  
QY 320 -----QRRESSRPVKPKKVDPSQOH-----PAPEKSKVSEQLKCCSGILKEMFAK 367  
DB 841 AESKPEQQAQKREPTPTPTPTDTSFQEDLIKFLPVPWEKLDKSED-----885  
QY 368 KHAAYAMPFKYVDVFEALGLHDYCDIKHPMDSTIKSLAREYDAQFAGDVRIMFS 427  
DB 886 -----AAPFRVPVDKLLNPDYHEIKRPMDELTVHKLYAGQYQAGQFCDDIWLMLD 940  
QY 428 NCYKYNPPDHEVVMARKLQDVF 450  
DB 941 NAWLYNRKNSKYKYGLKLESEM 963  
  
RESULT 12  
NFH\_RAT  
ID NFH\_RAT STANDARD; PRT; 831 AA.  
AC P16884; Q63368;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Neurofilament triplet H protein (200 kDa neurofilament protein)  
DE (Neurofilament heavy polypeptide) (NF-H) (Fragment).  
GN NEFH OR NFH.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=89065087; PubMed=3143606;  
RA Breen K.C., Robinson P.A., Wion D., Anderton B.H.;  
RT "Partial sequence of the rat heavy neurofilament polypeptide (NF-H).  
RT Identification of putative phosphorylation sites.";  
RL FEBS Lett. 241:213-218(1988).  
RN [2]  
RP SEQUENCE OF 37-831 FROM N.A.  
RX MEDLINE=88309090; PubMed=2457365;

RA Dautigny A., Pham-Dinh D., Roussel C., Felix J.M., Nussbaum J.L.,  
RT Jolles P.;  
RT "The large neurofilament subunit (NF-H) of the rat: cDNA cloning and  
RT in situ detection.";  
RL Biochem. Biophys. Res. Commun. 154:1099-1106(1988).  
RN [3]  
RP SEQUENCE OF 1-89 AND 243-313 FROM N.A.  
RX MEDLINE=87080760; PubMed=2878828;  
RA Robinson P.A., Wion D., Anderton B.H.;  
RT "Isolation of a cDNA for the rat heavy neurofilament polypeptide  
RT (NF-H).";  
RL FEBS Lett. 209:203-205(1986).  
RN [4]  
RP SEQUENCE OF 318-831 FROM N.A.  
RX MEDLINE=89184647; PubMed=2928342;  
RA Lieberburg I., Spalner N., Snyder S., Anderson J., Goldgaber D.,  
RT Smulowitz M., Carroll Z., Emanuel B.S., Breitner J., Rubin L.;  
RT "Cloning of a cDNA encoding the rat high molecular weight  
RT neurofilament peptide (NF-H): developmental and tissue expression in  
RT the rat, and mapping of its human homologue to chromosomes 1 and  
RT 22.";  
RL Proc. Natl. Acad. Sci. U.S.A. 86:2463-2467(1989).  
CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,  
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.  
CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT  
CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.  
CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFH IS  
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS  
CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF  
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE  
CC OF AXONAL CALIBER.  
CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING  
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H). THE  
CC LEVELS OF PHOSPHORYLATION BRING ALTERED DEVELOPMENTALLY AND  
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.  
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 783  
CC ONWARD AND IS LONGER DUE TO A FRAMESHIFT.  
  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
  
EMBL: M37227; AAA41693.1; ALT\_FRAME.  
DR EMBL: X13804; CAA32038.1; ALT\_FRAME.  
DR EMBL: M21964; AAA41695.1; -.  
DR EMBL: J04517; AAA41692.1; -.  
DR PIR: A30796; A30796.  
DR PIR: A25649; A25649.  
DR PIR: B25649; B25649.  
DR PIR: S02003; S02003.  
DR InterPro: IPR001684; IF.  
DR Pfam: PF00038; filament; 1.  
DR PROSITE: PS00226; IF; 1.  
KW Intermediate filament; Coiled coil; Neurone; Phosphorylation; Repeat.  
FT NON\_TER 1 1  
FT DOMAIN 276 641 51 X 3 AA TANDEM REPEATS OF K-S-P.  
FT CONFLICT 164 185 L -> I (IN REF. 2).  
FT CONFLICT 185 185 I -> S (IN REF. 2).  
FT CONFLICT 193 193 L -> T (IN REF. 2).  
FT CONFLICT 199 199 M -> T (IN REF. 2).  
FT CONFLICT 346 346 K -> N (IN REF. 1).  
FT CONFLICT 373 373 A -> V (IN REF. 1 AND 4).  
FT CONFLICT 482 482 G -> E (IN REF. 2 AND 4).  
FT CONFLICT 485 485 P -> S (IN REF. 2).  
FT CONFLICT 570 571 RK -> KE (IN REF. 2 AND 4).  
FT CONFLICT 591 591 P -> T (IN REF. 2 AND 4).  
FT CONFLICT 727 727 A -> V (IN REF. 4).  
FT CONFLICT 757 759 AAP -> GST (IN REF. 4).



(Neurofilament heavy polypeptide) (NF-H).  
NEFH OR NFH.  
Mus musculus (Mouse).  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
NCBI\_TaxID=10090;  
[1]  
SEQUENCE FROM N.A.  
MEDLINE=89121513; PubMed=3220257;  
Julien J.-P., Cote F., Beaudet L., Sidky M., Flavell D., Grosveld F.,  
Mushynski W.;  
"Sequence and structure of the mouse gene coding for the largest  
neurofilament subunit.";  
Gene 68:307-314(1988).  
[2]  
SEQUENCE FROM N.A.  
MEDLINE=89089138; PubMed=3145094;  
Sheldman P.S., Carden M.J., Lees J.F., Lazzarini R.A.;  
"The structure of the largest murine neurofilament protein (NF-H) as  
revealed by cDNA and genomic sequences";  
Brain Res. 464:217-231(1988).  
[3]  
SEQUENCE FROM N.A.  
STRAIN=Swiss Webster; TISSUE=Brain;  
Carden M.J.;  
Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
-!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,  
AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.  
NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT  
SUBSERVED BY THE TWO SMALLER NF PROTEINS.  
-!- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P. NFH IS  
PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS  
THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF  
INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE  
OF AXONAL CALIBER.  
-!- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING  
OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H). THE  
LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND  
COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.  
-!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
-!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS 534  
TO 716 AND IS SHORTER DUE TO FRAMESHIFTS.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
-----  
EMBL; M24496; AAA39813.1;  
EMBL; M23349; AAA39813.1; JOINED.  
EMBL; M24494; AAA39813.1; JOINED.  
EMBL; M24495; AAA39813.1; JOINED.  
EMBL; M35131; AAA39809.1; JOINED.  
EMBL; M35131; AAA39809.1; ALT\_FRAME.  
EMBL; Z31012; CA833229.1;  
PIR; JT0368; QFMSH.  
PIR; A43778; A43778.  
DR MGD; MGI:97309; Nfh.  
DR InterPro; IPR001664; IF.  
DR Pfam; PF00038; filament; 1.  
DR PROSITE; PS00226; IF; 1.  
KW Intermediate filament; Coiled coil; Neurone; Phosphorylation;  
Repeat.  
FT DOMAIN 1 97 HEAD.  
FT ROD. 408  
FT DOMAIN 409 1087 TAIL.  
FT DOMAIN 436 517 GLU-RICH (ACIDIC).  
FT DOMAIN 519 886 50 X 6 AA TANDEM REPEATS OF K-S-P-A-E-A.  
FT DOMAIN 887 1087 GLU/LYS-RICH.  
FT DOMAIN 98 129 COIL 1A.  
FT DOMAIN 130 141 LINKER 1.

DE (Neurofilament heavy polypeptide) (NF-H).  
GN NEFH OR NFH.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89121513; PubMed=3220257;  
RA Julien J.-P., Cote F., Beaudet L., Sidky M., Flavell D., Grosveld F.,  
RA Mushynski W.;  
RT "Sequence and structure of the mouse gene coding for the largest  
neurofilament subunit.";  
RL Gene 68:307-314(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89089138; PubMed=3145094;  
RA Sheldman P.S., Carden M.J., Lees J.F., Lazzarini R.A.;  
RT "The structure of the largest murine neurofilament protein (NF-H) as  
revealed by cDNA and genomic sequences";  
RL Brain Res. 464:217-231(1988).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Swiss Webster; TISSUE=Brain;  
RA Carden M.J.;  
CC Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,  
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.  
CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT  
CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.  
CC -!- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P. NFH IS  
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS  
CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF  
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE  
CC OF AXONAL CALIBER.  
CC -!- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING  
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H). THE  
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND  
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.  
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS 534  
CC TO 716 AND IS SHORTER DUE TO FRAMESHIFTS.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
-----  
EMBL; M24496; AAA39813.1;  
EMBL; M23349; AAA39813.1; JOINED.  
EMBL; M24494; AAA39813.1; JOINED.  
EMBL; M24495; AAA39813.1; JOINED.  
EMBL; M35131; AAA39809.1; JOINED.  
EMBL; M35131; AAA39809.1; ALT\_FRAME.  
EMBL; Z31012; CA833229.1;  
PIR; JT0368; QFMSH.  
PIR; A43778; A43778.  
DR MGD; MGI:97309; Nfh.  
DR InterPro; IPR001664; IF.  
DR Pfam; PF00038; filament; 1.  
DR PROSITE; PS00226; IF; 1.  
KW Intermediate filament; Coiled coil; Neurone; Phosphorylation;  
Repeat.  
FT DOMAIN 1 97 HEAD.  
FT ROD. 408  
FT DOMAIN 409 1087 TAIL.  
FT DOMAIN 436 517 GLU-RICH (ACIDIC).  
FT DOMAIN 519 886 50 X 6 AA TANDEM REPEATS OF K-S-P-A-E-A.  
FT DOMAIN 887 1087 GLU/LYS-RICH.  
FT DOMAIN 98 129 COIL 1A.  
FT DOMAIN 130 141 LINKER 1.

DE (Neurofilament heavy polypeptide) (NF-H).  
GN NEFH OR NFH.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89121513; PubMed=3220257;  
RA Julien J.-P., Cote F., Beaudet L., Sidky M., Flavell D., Grosveld F.,  
RA Mushynski W.;  
RT "Sequence and structure of the mouse gene coding for the largest  
neurofilament subunit.";  
RL Gene 68:307-314(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89089138; PubMed=3145094;  
RA Sheldman P.S., Carden M.J., Lees J.F., Lazzarini R.A.;  
RT "The structure of the largest murine neurofilament protein (NF-H) as  
revealed by cDNA and genomic sequences";  
RL Brain Res. 464:217-231(1988).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Swiss Webster; TISSUE=Brain;  
RA Carden M.J.;  
CC Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,  
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.  
CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT  
CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.  
CC -!- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P. NFH IS  
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS  
CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF  
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE  
CC OF AXONAL CALIBER.  
CC -!- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING  
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H). THE  
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND  
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.  
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS 534  
CC TO 716 AND IS SHORTER DUE TO FRAMESHIFTS.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
-----  
EMBL; M24496; AAA39813.1;  
EMBL; M23349; AAA39813.1; JOINED.  
EMBL; M24494; AAA39813.1; JOINED.  
EMBL; M24495; AAA39813.1; JOINED.  
EMBL; M35131; AAA39809.1; JOINED.  
EMBL; M35131; AAA39809.1; ALT\_FRAME.  
EMBL; Z31012; CA833229.1;  
PIR; JT0368; QFMSH.  
PIR; A43778; A43778.  
DR MGD; MGI:97309; Nfh.  
DR InterPro; IPR001664; IF.  
DR Pfam; PF00038; filament; 1.  
DR PROSITE; PS00226; IF; 1.  
KW Intermediate filament; Coiled coil; Neurone; Phosphorylation;  
Repeat.  
FT DOMAIN 1 97 HEAD.  
FT ROD. 408  
FT DOMAIN 409 1087 TAIL.  
FT DOMAIN 436 517 GLU-RICH (ACIDIC).  
FT DOMAIN 519 886 50 X 6 AA TANDEM REPEATS OF K-S-P-A-E-A.  
FT DOMAIN 887 1087 GLU/LYS-RICH.  
FT DOMAIN 98 129 COIL 1A.  
FT DOMAIN 130 141 LINKER 1.

DE (Neurofilament heavy polypeptide) (NF-H).  
GN NEFH OR NFH.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MED





[illegible][illegible]

RESULT 15
VRP1_YEAST
ID VRP1_
AC P3737
DT 01-00
DT 01-NO

```
Qy 54 NKPKROTNOLOYLLRVVLKTLWKHQFAPFPQOPVDVAVKL-----NLPDYKIKITPMD 106
Db 134 PIPNAPLS-----PAPVPSIPSSAPPIDIPSSAAPPIP 169
Qy 107 MGTIKKRLNNYYWNAOECIQDFNTMETNICYIYKPGDDIVLMAEAKFLQKINELPT 166
Db 170 I-----VPSSAPPPLSGASAPK-----VPQ 191
Qy 167 EETIMIVQAKGRGRGKETCTAKGVSTVP-----NTQASTPPQTQTPQPNP----- 215
Db 192 NRPHMPSVRPAHRSHQRKSSNISLPSVSAPPLPSASLPTHVSNPQAPPPTPTIGLDS 251
Qy 216 -----PP-----VOATPHFPFPAVTPDL 232
Db 252 KNIKPTDNVSPSEVPAGCLPLAEINARRSERGAVEGVSTKIOTENHKSPS----- 306
Qy 233 IVOTPVMTVVPPQLOTPPPVPPQPPPP-----APAPQPVQSHPPITIAAT 278
Db 307 --QPPLPSSAPIPTSHAPPLPPTAPPPPSLPNVTSAKKATSAPAPP-----PPLPAA- 359
Qy 279 PQPVTKKGVKRKADTTPTTTIDPIHEPPSL-PPEPKTKLQQRRESSRVKPKKDVDP 337
Db 360 -----MSSASTNSVKATEV--PPTLAPPLPNTTSVPPNKAASSMPAPPPPP----- 402
Qy 338 SQHPAPEKSKSVSEQLKCCSGILKEMFAKHAAYANPFYKPDVDEALGLHDYCDIHKP 397
Db 403 -----PPPGAFSTSSALSASSIPLAPL----- 425
Qy 398 MDMSTIKSLEAREYRDAQEFGADVRLMFSNCYKYNPPDHEVWAMARKLODVFEMRFKM 457
Db 426 -----PP----- 427
Qy 458 PDEPEEPVAVSPVPPPTKVA---PPSSSDSSSDSSDSDSDTDDSEERAQRLAEL 514
Db 428 ---PPPPSVATSVSPAPPPPTLTNKPSSASSKQSISSSSSSSAVTPGGP--LPFLAEI 482
Qy 515 QEOLKAVHEQLAALSQONKPKKKEKKKKKKKEEVEENKSKAKEPPPKTK 574
Db 483 QKK-----RDDRFFVGGDTGYTTQDKQEDVIGSSKDDNVSPISPSIN-----PPQSS 532
Qy 575 KNSS-----NSNVSKK-----EPAPMKSKPPPTVESEED 605
Db 533 QNGMSFELDETESKLHKQTSNAFNAPPHTDAMAPPLPSPAPPPPTTSLPTPTASGDDHT 592
Qy 606 KCKPMSTEERKQLSLDINK-----LPG 627
Db 593 -----NDRKETVLGMKKAKAPALPG 612
```

Search completed: February 25, 2003, 13:21:03  
Job time : 23 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2003, 13:17:31 ; Search time 41 Seconds  
(without alignments)  
3457.571 Million cell updates/sec

Title: US-09-700-590A-22  
Perfect score: 3639  
Sequence: 1 MSASGPGTRLNLPVMDG.....ALCHLLFAEKEKTFKLRLKLM 688

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_virus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3531.5	97.0	731	4	O60433
2	3397	93.3	723	11	Q8VHF7
3	3392	93.2	1400	11	Q9ESU6
4	3392	93.2	1400	11	Q8VHF8
5	1915	52.6	726	11	Q8J125
6	1901	52.2	798	11	Q88411
7	1901	52.2	798	11	O54795
8	1887	51.9	801	4	Q969U4
9	1876.5	51.6	729	13	Q90971
10	1757	48.3	701	13	Q8QF77
11	1690.5	46.5	814	13	Q8U0M2
12	1680	46.2	664	5	Q8T775
13	1678	46.1	732	5	Q9GU61
14	1540	42.3	956	11	Q91Y44
15	1501.5	41.3	947	4	O14789
16	1401	38.5	503	11	Q99PC5

17	1339.5	36.8	1937	5	Q9W3L3
18	763.5	21.0	851	5	Q9S780
19	608.5	16.7	1087	5	Q20948
20	600.5	16.5	1250	5	Q20947
21	561	15.4	578	3	Q9Y7N0
22	531.5	14.6	727	3	Q9HGP4
23	498.5	13.7	249	13	Q91686
24	495.5	13.6	638	3	Q07442
25	430	11.8	513	5	Q9VCG6
26	375	10.3	507	10	Q9LHY2
27	368.5	10.1	361	3	P78808
28	365.5	10.0	128	11	O35692
29	351.5	9.7	370	5	Q8SR93
30	344	9.5	643	10	Q9LSL2
31	339.5	9.3	461	10	Q9S7T1
32	339	9.3	678	10	Q9LXA7
33	337	9.3	602	10	Q9AR00
34	323.5	8.9	766	10	Q9LNC4
35	322.5	8.9	247	5	Q8T3Z6
36	318.5	8.8	268	5	Q9WIS2
37	318	8.7	440	10	Q9LMU8
38	315	8.7	487	10	Q93ZU2
39	314	8.6	2429	11	Q8QZV8
40	303.5	8.3	818	10	Q9LK27
41	300.5	8.3	3080	5	Q9VRY3
42	297.5	8.2	374	5	O76561
43	285	7.8	1188	10	Q41805
44	282.5	7.8	1235	4	Q9H0E9
45	282	7.7	369	10	Q9FT54

#### ALIGNMENTS

#### RESULT 1

O60433

ID O60433 PRELIMINARY; PRT; 731 AA.

AC O60433; (TREMBLrel. 07, Created)

DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE R31546\_1 (Fragment).

GN HUNK1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Lamerdin J.E., McCready P.M., Skowronski E., Adamson A.W.,

RA Burkhart-Schultz K.J., Gordon L., Kyle A., Ramirez M., Stillwagen S.,

RA Phan H., Velasco N., Do L., Regala W., Terry A., Barnes J.,

RA Danganan L., Poundstone P., Christensen M., Georgescu A., Avila J.,

RA Liu S., Attix C., Andreise T., Trankheim M., Amico-Keller G.,

RA Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,

RA Krommiller B., Arellano A., Montgomery M., Ow D., Nolan M., Trong S.,

RT Kobayashi A., Olsen A.S., Carrano A.V.;

RT "Sequence analysis of an ~1.5 MB OLF cluster in 19p13.1.1.;

RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; AC004798; AAC27978.1; -

DR HSSP; Q92831; 1B91.

DR InterPro; IPR001487; Bromodomain.

DR Pfam; PF00439; bromodomain; 2.

DR PRINTS; PR00503; BROMODOMAIN.

DR SMART; SM00297; BROMO; 2.

DR PROSITE; PS00633; BROMODOMAIN\_1; 1.

DR PROSITE; PS50014; BROMODOMAIN\_2; 2.

FT NON\_TER

SQ SEQUENCE 731 AA; 81383 MW; F56C3A02F1A26F65 CRC64;

Query Match 97.0%; Score 3531.5; DB 4; Length 731;

Best Local Similarity 98.0%; Pred. No. 2.1e-200;

Matches 672; Conservative 0; Mismatches 5; Indels 9; Gaps 1;

```
QY 1 MSASGPGTRLNRLNLPVMDGLTSSOMSTTQAQAQOPANAASTNPPPETSNPNKPKROT 60
DB 12 MSASGPGTRLNRLNLPVMDGLTSSOMSTTQAQAQOPANAASTNPPPETSNPNKPKROT 71
QY 61 NOLQYLLRVLLKTLWKHGFAMFPQOPVDAVKLNLPDYKIIKTPMDMGTIKKRLNNYIW 120
DB 72 NOLQYLLRVLLKTLWKHGFAMFPQOPVDAVKLNLPDYKIIKTPMDMGTIKKRLNNYIW 131
QY 121 NAEQIQDFTNFTNCYIYNKPGDDIVLMAEALFKLFLQKINELPTEETEIMIVQAKGRG 180
DB 132 NAEQIQDFTNFTNCYIYNKPGDDIVLMAEALFKLFLQKINELPTEETEIMIVQAKGRG 191
QY 181 RGRKETGTAQGVSTVPNTTQASTPQOTQOPNPPVQATPHPPAVTDLIVQTPVMT 240
DB 192 RGRKETGTAQGVSTVPNTTQASTPQOTQOPNPPVQATPHPPAVTDLIVQTPVMT 251
QY 241 VVPPQPLQTPPPVPPQOPPPAPAPQVQSHPPPIIAATPQVKTGKVKRKADTTPTTI 300
DB 252 VVPPQPLQTPPPVPPQOPPPAPAPQVQSHPPPIIAATPQVKTGKVKRKADTTPTTI 311
QY 301 DPIHEPPSLPPEPKTKLQGRRESSRPVKKPKDVDSQQHPAPEKSKVSEQLKCCSGI 360
DB 312 DPIHEPPSLPPEPKTKLQGRRESSRPVKKPKDVDSQQHPAPEKSKVSEQLKCCSGI 371
QY 361 LKEMFAKHAAYAMPFYKPDVEALGLHDYCDIHKHPMDMTIKSKLEAREYRDAQERGA 420
DB 372 LKEMFAKHAAYAMPFYKPDVEALGLHDYCDIHKHPMDMTIKSKLEAREYRDAQERGA 431
QY 421 DVLRFMSNCYKYNPDHEVAMARKLDQVFEMFAKMPDEPEEPVAVSSPAVPPPTKV 480
DB 432 DVLRFMSNCYKYNPDHEVAMARKLDQVFEMFAKMPDEPEEPVAVSSPAVPPPTKV 491
QY 481 APPSSSSSSSSSSSSSSSSSSSSSEERAQRLAEQLKAVHEQLAALSQQQNPKPKKE 540
DB 492 APPSSSSSSSSSSSSSSSSSSSSSEERAQRLAEQLKAVHEQLAALSQQQNPKPKKE 551
QY 541 KDKKKEKKHKKKEVEENKSKAKPPPKTKKNSSNSNVSKKEPAPMKSPPPPYE 600
DB 552 KDKKKEKKHKKKEVEENKSKAKPPPKTKKNSSNSNVSKKEPAPMKSPPPPYE 611
QY 601 SEEDKCKPMSEYERKQLSLDINKLPGKLGKRVVHIIQSRPSLKNPNPDEIDFETLK 660
DB 612 SEEDKCKPMSEYERKQLSLDINKLPGKLGKRVVHIIQSRPSLKNPNPDEIDFETLK 671
QY 661 PSTLRELKALCHLLFAEKEKTFKLK 686
DB 672 PSTLREL-----ERYVTSCLRK 688
```

## RESULT 2

```
Q8VHF7 ID Q8VHF7 PRELIMINARY; PRT; 723 AA.
AC Q8VHF7;
DT 01-MAR-2002 (TremBLrel. 20, Created)
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Bromodomain-containing protein BRD4 short variant.
GN BRD4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Houselstein D., Bullock S.L., Lynch D.E., Grigorieva E.F.,
RA Wilson V.A., Bedington R.S.P.;
RT "Growth and early post implantation defects in mice mutant for the
RT bromodomain-containing protein Brd4."
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF461396; AAL67834.1; -.
DR InterPro; IPR001487; Bromodomain.
DR Pfam; PF00439; bromodomain; 2.
```

```
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 2.
DR PROSITE; PS00633; BROMODOMAIN_1; UNKNOWN_1.
DR PROSITE; PS0014; BROMODOMAIN_2; 2.
SQ SEQUENCE 723 AA; 80635 MW; 738F2AE5F58A56BC CRC64;

Query Match 93.3%; Score 3397; DB 11; Length 723;
Best Local Similarity 94.3%; Pred. No. 1.7e-192;
Matches 648; Conservative 5; Mismatches 24; Indels 10; Gaps 2;

QY 1 MSASGPGTRLNRLNLPVMDGLTSSOMSTTQAQAQOPANAASTNPPPETSNPNKPKROT 60
DB 12 MSASGPGTRLNRLNLPVMDGLTSSOMSTTQAQAQOPANAASTNPPPETSNPNKPKROT 60
QY 61 NOLQYLLRVLLKTLWKHGFAMFPQOPVDAVKLNLPDYKIIKTPMDMGTIKKRLNNYIW 120
DB 61 NOLQYLLRVLLKTLWKHGFAMFPQOPVDAVKLNLPDYKIIKTPMDMGTIKKRLNNYIW 120
QY 121 NAEQIQDFTNFTNCYIYNKPGDDIVLMAEALFKLFLQKINELPTEETEIMIVQAKGRG 180
DB 121 NAEQIQDFTNFTNCYIYNKPGDDIVLMAEALFKLFLQKINELPTEETEIMIVQAKGRG 180
QY 181 RGRKETGTAQGVSTVPNTTQASTPQOTQOPNPPVQATPHPPAVTDLIVQTPVMT 239
DB 181 RGRKETGTAQGVSTVPNTTQASTPQOTQOPNPPVQATPHPPAVTDLIVQTPVMT 240
QY 240 TVVPPQPLQTPPPVPPQOPPPAPAPQVQSHPPPIIAATPQVKTGKVKRKADTTPTT 299
DB 241 TVVPPQPLQTPPPVPPQOPPPAPAPQVQSHPPPIIAATPQVKTGKVKRKADTTPTT 300
QY 300 IDPIHEPPSLPPEPKTKLQGRRESSRPVKKPKDVDSQQHPAPEKSKVSEQLKCCSG 359
DB 301 IDPIHEPPSLPPEPKTKLQGRRESSRPVKKPKDVDSQQHPAPEKSKVSEQLKCCSG 360
QY 360 ILKEMFAKHAAYAMPFYKPDVEALGLHDYCDIHKHPMDMTIKSKLEAREYRDAQERGA 419
DB 361 ILKEMFAKHAAYAMPFYKPDVEALGLHDYCDIHKHPMDMTIKSKLEAREYRDAQERGA 420
QY 420 ADVRLMFNSCYKYNPDHEVAMARKLDQVFEMFAKMPDEPEEPVAVSSPAVPPPTKV 475
DB 421 ADVRLMFNSCYKYNPDHEVAMARKLDQVFEMFAKMPDEPEEPVAVSSPAVPPPTKV 480
QY 480 VAPSSSSSSSSSSSSSSSSSSSSSEERAQRLAEQLKAVHEQLAALSQQQNPKPKK 539
DB 481 VAPSSSSSSSSSSSSSSSSSSSSSEERAQRLAEQLKAVHEQLAALSQQQNPKPKK 540
QY 540 EKDKEKKHKKKEVEENKSKAKPPPKTKKNSSNSNVSKKEPAPMKSPPPPYE 599
DB 541 EKDKEKKHKKKEVEENKSKAKPPPKTKKNSSNSNVSKKEPAPMKSPPPPYE 600
QY 600 ESEEDKCKPMSEYERKQLSLDINKLPGKLGKRVVHIIQSRPSLKNPNPDEIDFETLK 659
DB 601 ESEEDKCKPMSEYERKQLSLDINKLPGKLGKRVVHIIQSRPSLKNPNPDEIDFETLK 660
QY 660 KPSLRELKALCHLLFAEKEKTFKLK 686
DB 661 KPSLREL-----ERYVTSCLRK 678
```

## RESULT 3

```
Q9ESU6 ID Q9ESU6 PRELIMINARY; PRT; 1400 AA.
AC Q9ESU6;
DT 01-MAR-2001 (TremBLrel. 16, Created)
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Cell proliferation related protein CAP.
GN BRD4 OR CAP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RX MEDLINE-20396330; PubMed-10938129;
RA Dey A., Ellenberg J., Farina A., Coleman A.E., Maruyama T.,
RT Sciortino S., Lippincott-Schwartz J., Ozato K.;
RT "A bromodomain protein, MCPAP, associates with mitotic chromosomes and
RL affects G(2)-to-M transition."
RL Mol. Cell. Biol. 20:6537-6549(2000).
DR EMBL; AF273217; AAG02191.1; -
DR HSSP; Q92831; 1891.
DR MGD; MGI:1888520; Brd4.
DR InterPro; IPR001487; Bromodomain.
DR Pfam; PF00439; bromodomain; 2.
DR PRINTS; PR00503; BROMODOMAIN.
DR PRINTS; PR01217; PRICHEXTENS.
DR SMART; SM00297; BROMO; 2.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS50014; BROMODOMAIN_2; 2.
SQ SEQUENCE 1400 AA; 155923 MW; 9902BFF7B00ADB59 CRC64;

Query Match 93.2%; Score 3392; DB 11; Length 1400;
Best Local Similarity 94.2%; Pred. No. 6.8e-192;
Matches 647; Conservative 5; Mismatches 25; Indels 10; Gaps 2;

QY 1 MSAESGPGTRLRLNLPVMDGLETSMSTTQAAQAPQAPANAASNNPPPETSNPKPKROT 60
DB 1 MSTESGPGTRLRLNLPVMDGLETSMSTTQAAQAPQAPANAASNNPPPETSNPKPKROT 60
QY 61 NQLQYLLRVVLKTLWKHQFAWPFQOPVDVAVKLNLPDYKIIKTPMDMGTKKRLNNYYW 120
DB 61 NQLQYLLRVVLKTLWKHQFAWPFQOPVDVAVKLNLPDYKIIKTPMDMGTKKRLNNYYW 120
QY 121 NAQECIDQNTMTNCTNYYKPGDDIVLMAEALFKLQKINELPTEETETIMIVQAKRG 180
DB 121 NAQECIDQNTMTNCTNYYKPGDDIVLMAEALFKLQKINELPTEETETIMIVQAKRG 180
QY 181 RGRKETGAKPGVSTVNTTQASTPQTPQPN-PPVQATPHPPFAVTPDLIVQTPVM 239
DB 181 RGRKETGAKPGVSTVNTTQASTPQTPQPN-PPVQATPHPPFAVTPDLIVQTPVM 239
QY 240 TVVPPQPLQTPSPVPPQPPAPAPQVQSHPPPIAATPQPVKTKGVRKADTTPTT 299
DB 240 TVVPPQPLQTPSPVPPQPPAPAPQVQSHPPPIAATPQPVKTKGVRKADTTPTT 299
QY 300 IDPIHEPPSLPPEPKTKLQORRESSRPVKPKDVPDSQOHPAPEKSKVSSEOLKCCSG 359
DB 300 IDPIHEPPSLPPEPKTKLQORRESSRPVKPKDVPDSQOHPAPEKSKVSSEOLKCCSG 359
QY 360 ILKEMFAKKAAYAWPFYKPDVEALGLHDYCDIILKHPMDMTIKSLAREYDAQEF 419
DB 360 ILKEMFAKKAAYAWPFYKPDVEALGLHDYCDIILKHPMDMTIKSLAREYDAQEF 419
QY 420 ADVRLMFSNCKYKYNPDHEVYVAMARKLDQVFEMFAKMPDEPEPVVAVSSPAVPTKV 479
DB 420 ADVRLMFSNCKYKYNPDHEVYVAMARKLDQVFEMFAKMPDEPEPVVAVSSPAVPTKV 479
QY 480 VAPSSSDSSSDSSSDSDTDDSEERAQRLAEQLKAVHEQLAALSQOONKPKKK 539
DB 480 VAPSSSDSSSDSSSDSDTDDSEERAQRLAEQLKAVHEQLAALSQOONKPKKK 539
QY 540 EKDKKKKKKKKKKEEVENKSKAKEPPKTKKNNSSNSVSKKEPAPMKSKPPPT 599
DB 540 EKDKKKKKKKKKKEEVENKSKAKEPPKTKKNNSSNSVSKKEPAPMKSKPPPT 599
QY 600 ESEEDCKPMSYEKQLSLDINKLPGEKLGVRVHIIQSRPSLKNPNDEIDFETL 659
DB 600 ESEEDCKPMSYEKQLSLDINKLPGEKLGVRVHIIQSRPSLKNPNDEIDFETL 659
QY 660 KPSTLRELGAHLALFAEKEKTKLRK 686
DB 660 KPSTLRELGAHLALFAEKEKTKLRK 686
```

```
RESULT 4
Q8VHF8 PRELIMINARY; PRT; 1400 AA.
ID Q8VHF8;
AC Q8VHF8;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Bromodomain-containing protein BRD4 long variant.
GN BRD4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Houzelstein D., Bullock S.L., Lynch D.E., Grigorieva E.F.,
RA Wilson V.A., Beddington R.S.P.;
RT "Growth and early post implantation defects in mice mutant for the
RT bromodomain-containing protein Brd4."
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF461395; AAL67833.1; -
DR InterPro; IPR001487; Bromodomain.
DR Pfam; PF00439; bromodomain; 2.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 2.
DR PROSITE; PS00633; BROMODOMAIN_1; UNKNOWN_1.
DR PROSITE; PS50014; BROMODOMAIN_2; 2.
SQ SEQUENCE 1400 AA; 155925 MW; 89952B9E75501BC4 CRC64;

Query Match 93.2%; Score 3392; DB 11; Length 1400;
Best Local Similarity 94.2%; Pred. No. 6.8e-192;
Matches 647; Conservative 5; Mismatches 25; Indels 10; Gaps 2;

QY 1 MSAESGPGTRLRLNLPVMDGLETSMSTTQAAQAPQAPANAASNNPPPETSNPKPKROT 60
DB 1 MSTESGPGTRLRLNLPVMDGLETSMSTTQAAQAPQAPANAASNNPPPETSNPKPKROT 60
QY 61 NQLQYLLRVVLKTLWKHQFAWPFQOPVDVAVKLNLPDYKIIKTPMDMGTKKRLNNYYW 120
DB 61 NQLQYLLRVVLKTLWKHQFAWPFQOPVDVAVKLNLPDYKIIKTPMDMGTKKRLNNYYW 120
QY 121 NAQECIDQNTMTNCTNYYKPGDDIVLMAEALFKLQKINELPTEETETIMIVQAKRG 180
DB 121 NAQECIDQNTMTNCTNYYKPGDDIVLMAEALFKLQKINELPTEETETIMIVQAKRG 180
QY 181 RGRKETGAKPGVSTVNTTQASTPQTPQPN-PPVQATPHPPFAVTPDLIVQTPVM 239
DB 181 RGRKETGAKPGVSTVNTTQASTPQTPQPN-PPVQATPHPPFAVTPDLIVQTPVM 239
QY 240 TVVPPQPLQTPSPVPPQPPAPAPQVQSHPPPIAATPQPVKTKGVRKADTTPTT 299
DB 240 TVVPPQPLQTPSPVPPQPPAPAPQVQSHPPPIAATPQPVKTKGVRKADTTPTT 299
QY 300 IDPIHEPPSLPPEPKTKLQORRESSRPVKPKDVPDSQOHPAPEKSKVSSEOLKCCSG 359
DB 300 IDPIHEPPSLPPEPKTKLQORRESSRPVKPKDVPDSQOHPAPEKSKVSSEOLKCCSG 359
QY 360 ILKEMFAKKAAYAWPFYKPDVEALGLHDYCDIILKHPMDMTIKSLAREYDAQEF 419
DB 360 ILKEMFAKKAAYAWPFYKPDVEALGLHDYCDIILKHPMDMTIKSLAREYDAQEF 419
QY 420 ADVRLMFSNCKYKYNPDHEVYVAMARKLDQVFEMFAKMPDEPEPVVAVSSPAVPTKV 479
DB 420 ADVRLMFSNCKYKYNPDHEVYVAMARKLDQVFEMFAKMPDEPEPVVAVSSPAVPTKV 479
QY 480 VAPSSSDSSSDSSSDSDTDDSEERAQRLAEQLKAVHEQLAALSQOONKPKKK 539
DB 480 VAPSSSDSSSDSSSDSDTDDSEERAQRLAEQLKAVHEQLAALSQOONKPKKK 539
QY 540 EKDKKKKKKKKKKEEVENKSKAKEPPKTKKNNSSNSVSKKEPAPMKSKPPPT 599
DB 540 EKDKKKKKKKKKKEEVENKSKAKEPPKTKKNNSSNSVSKKEPAPMKSKPPPT 599
QY 541 EKDKKKKKKKKKKEEVENKSKAKEPPKTKKNNSSNSVSKKEPAPMKSKPPPT 600
DB 541 EKDKKKKKKKKKKEEVENKSKAKEPPKTKKNNSSNSVSKKEPAPMKSKPPPT 600
```

```
QY 600 ESEEDCKPMSYEEKRQLSLDINKLPGLKGRVHVHIIQSREPSLKNSNPDBIEIDFETL 659
|||||
DB 601 ESEEDCKPMSYEEKRQLSLDINKLPGLKGRVHVHIIQSREPSLKNSNPDBIEIDFETL 660
|||||
QY 660 KPSTRELKALCHLLFAEKEKTEFLRK 686
|||||
DB 661 KPSTREL-----ERYVTSLCKR 678
|||||

RESULT 5
ID Q9J125 PRELIMINARY; PRT; 726 AA.
AC Q9J125;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Bromodomain-containing FSH-like protein FSRG2.
GN BRD3 OR 2410084F24RIK OR FSRG2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Shang E., Wolgemuth D.J.;
RT "Cloning and expression pattern of Fsr2, a putative murine
RT bromodomain-containing homolog of the Drosophila gene female sterile
RT homeotic.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF269193; AAF78072.1; -.
DR HSSP; Q92831; IB91.
DR MGD; MGI:1914632; Brd3.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR002965; P-rich.extensn.
DR Pfam; PF00439; bromodomain; 2.
DR PRINTS; PR00503; BROMODOMAIN.
DR PRINTS; PR01217; PRICHEXTENS.
DR SMART; SM00297; BROMO; 2.
DR PROSITE; PS00633; BROMODOMAIN_1; 2.
DR PROSITE; PS0014; BROMODOMAIN_2; 2.
SQ SEQUENCE 726 AA; 79745 MW; 7AB3BADAD38A78F4 CRC64;

Query Match 52.6%; Score 1915; DB 11; Length 726;
Best Local Similarity 60.2%; Pred. No. 3.6e-105;
Matches 401; Conservative 64; Mismatches 141; Indels 60; Gaps 16;

QY 26 MSTTQAQAPQAPANAASNPPTTSNPKPKRQTNQLQYLLRVVLTWLKHQFAPFPQ 85
|||||
DB 1 MSTTAAATGIPAVCPVNPPTPEVSNFSGKGRKTNQLQYQWNVVVTWLKHQFAPFPYQ 60
|||||
QY 86 PVDAYKLNLPDYKIITPTMDMTGKIKRLNNYWNQAQECIQDFTMTNCTYIYNKPGDD 145
|||||
DB 61 PVDAILKLNLPDYHKIINKPMDMGTIKRLNNYWNQAQECIQDFTMTNCTYIYNKPTDD 120
|||||
QY 146 IYVMAEALFKLQINELPTEETIMIVQAKGR-----GRKETGAKGVSTVNTQTQ 201
|||||
DB 121 IYVMAEALFKLQINELPTEETIMIVQAKGR-----GRKETGAKGVSTVNTQTQ 178
|||||
QY 202 ASTPTQOTPPQN-PPVQAATPHPPAVTPDLIVOTPVYVPPQPLQTPPPVPPQPP 260
|||||
DB 179 A-----TPQNIPTVTSQP-----VIAATPVPTIT-----ANVTSVVPP 214
|||||
QY 261 PAPAPQVQSHPPITAAATPOPVKTKKGVKRRADTTTPT-----TIDPIHEPPSLPPEKTT 316
|||||
DB 215 PAAPPPPATPIPVVPPTP-PVVKKGVKRRADTTTPTTSAITSRSSEPPPL-SEPKQA 272
|||||
QY 317 KLGQRRES-SRPVKPKDVPDSQHPAPEKSSKVSQOLKCCSGILKEMFAKKAAYAMP 375
|||||
DB 273 KVARREGGRRIPKPPKKLLEDGEYQPHAGKKGKIKSEHLRCHDSILREMLSKKHAAYAMP 332
|||||
QY 376 FYKPVVDVALGLHDYCDIIKHPMDMTSTIKSLAEAREYDAQEFADVRLMFSNCKYKNPP 435
|||||
```

```
DB 333 FYKPVDAEALHLDYHDIHKHPMDLSTVYKRMDSREYDAQFAADIRLMFSNCKYKNPP 392
|||||
QY 436 DHEVYVAMARKLQDVPFMRFAKMPDEPVPVAVSSPVPVPPPTKVAVPSSSSSSSSSS 495
|||||
DB 393 DHEVYVAMARKLQDVPFMRFAKMPDEPME-----APALPAPTAPIVSKGASRSSESS 446
|||||
QY 496 SDSSTDDSEERAQRLAEQLQKAVHQBOLAALSQPOQNKP-KKKEKDKKK-----547
|||||
DB 447 SDGSSDSEERATRLAEQLQKAVHQBOLAALSQAPVYKPKKKKKKKKKKKDKDKD 506
|||||
QY 548 KEKHKKKEEVEENKSK-----AKEPPPK--TKKNNSNSNVSKKEPAPMKSKPPPTYES 601
|||||
DB 507 KEKEKHKAKSEEEKAKAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQA 564
|||||
QY 602 EEDCKPMSYEEKRQLSLDINKLPGLKGRVHVHIIQSREPSLKNSNPDBIEIDFETLKP 661
|||||
DB 565 EEEEGPLNSYDEKQLSLDINKLPGLKGRVHVHIIQSREPSLKNSNPDBIEIDFETLKP 624
|||||
QY 662 STLREL 667
|||||
DB 625 TTLREL 630
|||||

RESULT 6
ID O88411 PRELIMINARY; PRT; 798 AA.
AC O88411;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Female sterile homeotic-related protein Frg-1.
GN BRD2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Rhee K., Brunori M., Besset V., Wolgemuth D.J.;
RT "Expression and potential role of Frg-1, a putative murine
RT bromodomain-containing homolog of the Drosophila gene female sterile
RT homeotic.";
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF045462; AAC24810.1; -.
DR HSSP; Q92831; IB91.
DR MGD; MGI:99495; Brd2.
DR InterPro; IPR001487; Bromodomain.
DR Pfam; PF00439; bromodomain; 2.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 2.
DR PROSITE; PS00633; BROMODOMAIN_1; 2.
DR PROSITE; PS0014; BROMODOMAIN_2; 2.
SQ SEQUENCE 798 AA; 88063 MW; A9942517CF15B7A1 CRC64;

Query Match 52.2%; Score 1901; DB 11; Length 798;
Best Local Similarity 54.9%; Pred. No. 2.7e-104;
Matches 403; Conservative 86; Mismatches 151; Indels 94; Gaps 20;

QY 5 SGPGTRLNLPVMDGLTSMSTTQAQAPQAPANAASNPPTTSNPKPKRQTNQLQ 64
|||||
DB 26 AAGPKRIKPSLLYEGFESPTMSVPA-----LQAPANPPPPVSNPKPKRQTNQLQ 79
|||||
QY 65 YLLRVVLTWLKHQFAPFPQVDAVKLNLPDYKIITPTMDMTGKIKRLNNYWNQAQ 124
|||||
DB 80 YLHKVVMKALWKHQFAPFPQVDAVKLGLPDYHKIIPQMDMTGKIKRLNNYWNQAQ 139
|||||
QY 125 CIQDFTMTNCTYIYNKPGDDIVLMAEALFKLQINELPTEETIMIVQAKG---RG- 180
|||||
DB 140 CMQDFTMTNCTYIYNKPTDDIVLMAQTLKIFLQKVASMPQEQEQLVVTIPKNSHKRG 199
|||||
QY 181 -----RGKETGTAKPVGVSTVTPNTQASTPTQTPQPPVPPVQATPPFPAPTDLIVQ 235
|||||
DB 200 KLAALQGSITSASHQVAVSSVSHYALYTPPEI-----PTVLNIPHP-----SVIS 246
|||||
```



```
QY 236 TPVMTVVPQIQTPTPPVPQPPAPAPQPVQSHPPPIAATPQPVKTKGKVRKADTT 295
DB 247 SPLKSLH-----SAGPPLLAWSAAPP-----QPLAKKKGKVRKADTT 285
QY 296 TPTTIDPIHEP--PSLPP--EPKTKL--GORRESSRPVKPKKQVDPDSQOHPAPEKSSK 349
DB 286 TPTP-TAILAPGSPASPPGSLPEKAAARLPPMRRESGRPKPPKRLDPSQOQHSSKKGK 344
QY 350 VSEQLKCCSGILKEMFAKKAHAAYAWPFYKPDVVEALGLHDYCDIHKHPMDMTIKSKLEA 409
DB 345 LSEQLKHCNGILKELLSKKAHAAYAWPFYKPDVVEALGLHDYCDIHKHPMDLSTVKRKME 404
QY 410 REYRDAQEFAGADVRLMFSNCYKYNPPDHVWAMARKLDQVFMFAKMPDEPEEPVAVS 469
DB 405 RYRDAQEFAGADVRLMFSNCYKYNPPDHVWAMARKLDQVFEFYAKMPDEPLEGPLV 464
QY 470 SPAVPPP-TKVVAPPSSSDSSSS-----SDSDSTDDSEERAOHLAEQ 517
DB 465 STALPGLTKSSSESSSESSSESSSESSSESSSESSSESSSESSSEERAHRLAEQ 524
QY 518 LKAVHEQLAALSOPOONKPKKKEKKEKK-----EKHKKEVEENKKS-KAKEPP-PK 571
DB 525 LRVAHEQLAALSOQPIKPKRK-REKKEKKKKAKEHGRIGIDEDDKGPRAPRPPQPK 583
QY 572 KTKKNSSNSNV-----SKKEPAPMKSKPPPT-----YESEEDCKCKPMSE 613
DB 584 KSKKAGGGSNATTLSPHGFSGSSNKLPPKSKQKAPPVLTGYDSEESRPMSE 643
QY 614 EKRLSLDKLPGKGLGRVWHIIQSRPSLKNPNDETEIDFETLKPSTLRELGA-LCH 672
DB 644 EKRLSLDKLPGKGLGRVWHIIQAREPSLRDSNPPEEIEIDFETLKPSTLRELYVLS 703
QY 673 LLFAEEKETFKLRK 686
DB 704 CLRKKPKPYTIRK 717

RESULT 7
OS4795 PRELIMINARY; PRT; 798 AA.
AC OS4795;
DT 01-JUN-1998 (TREMREL. 06, Created)
DT 01-NOV-1998 (TREMREL. 08, Last sequence update)
DT 01-JUN-2002 (TREMREL. 21, Last annotation update)
DE RING3 protein.
GN BRD2 OR FSRG1 OR RING3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Thorpe K.L., Beck S.;
RL Immunogenetics 0:0-0(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129SVJ;
RA Taniguchi Y., Matsuzaka Y., Fujimoto H., Miyado K., Kohda A.,
RA Okumura K., Kimura M., Inoko H.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 47-549 FROM N.A.
RC STRAIN=INBRED CD-1; TISSUE=TESTIS;
RA Taniguchi Y., Matsuzaka Y., Fujimoto H., Miyado K., Kohda A.,
RA Okumura K., Kimura M., Inoko H.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=129SVJ;
RA Rowen L., Qin S., Madan A., Loretz C., James R., Dors M., Mix L.,
RA Hall J., Lasky S., Hood L.;
RT "Sequence of the mouse major histocompatibility locus class II
```

RT region. ";  
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL009226; CAA15818.1; -;  
DR EMBL; AL009226; CAA15819.1; -;  
DR EMBL; DB9801; BAA25416.1; -;  
DR EMBL; AB010248; BAA24379.1; -;  
DR EMBL; AB010247; BAA24378.1; -;  
DR EMBL; AB010246; BAA24377.1; -;  
DR EMBL; AF100956; AAC69907.1; -;  
DR HSSP; Q92831; I891.  
DR MGD; MGI:99495; Brd2.  
DR InterPro; IPR001487; Bromodomain.  
DR Pfam; PF00439; bromodomain; 2.  
DR PRINTS; PR00503; BROMODOMAIN.  
DR SMART; SM00297; BROMO; 2.  
DR PROSITE; PS00633; BROMODOMAIN\_1; 2.  
DR PROSITE; PS00633; BROMODOMAIN\_2; 2.  
SQ SEQUENCE 798 AA; 88066 MW; 08DD57FBF1385E96 CRC64;

Query Match 52.2%; Score 1901; DB 11; Length 798;

Best Local Similarity 54.9%; Pred. No. 2.7e-104;

Matches 403; Conservative 86; Mismatches 151; Indels 94; Gaps 20;

QY 5 SGPGLRLNLPVMDGLETSONSTTOAQOQOPANAASTNPPPETSNPNKPKROTNOIQ 64  
DB 26 AAPKKIRKPSLYEGFESPTMASVPA-----LQLAPANPPPEVSNPKKGRVTNQLQ 79  
QY 65 YLLRVLKTIWKHQFAWPFQPVDAVKLNPDIYKIKTPMDMGTIKKRLNNYNNAAOE 124  
DB 80 YLHKVYKALWKHQFAWPFQPVDAVKLGLPDYHKIIPKQMDGTIKRRLNNYNNAAOE 139  
QY 125 CIQDFNTMTNCTYIYNKPGDDIVLMAEALKEFLQKINELPTEETIMIVQAKG---RG- 180  
DB 140 CMQDFNTMTNCTYIYNKPTDDIVLMAQTLEKIFLOKVASMPQEEQLVVTIPKNSHKGA 199  
QY 181 -----RGRKETGAKPGVSTVNTTOASTPQQTQTPQNPVQOATPHPPATVPLIVQ 235  
DB 200 KLAALQGSITSAHQVPAVSVSHTALYTPPPEI-----PTTVLNIHPH-----SVIS 246  
QY 236 TPVMTVVPQIQTPTPPVPQPPAPAPQPVQSHPPPIAATPQPVKTKGKVRKADTT 295  
DB 247 SPLKSLH-----SAGPPLLAWSAAPP-----QPLAKKKGKVRKADTT 285  
QY 296 TPTTIDPIHEP--PSLPP--EPKTKL--GORRESSRPVKPKKQVDPDSQOHPAPEKSSK 349  
DB 286 TPTP-TAILAPGSPASPPGSLPEKAAARLPPMRRESGRPKPPKRLDPSQOQHSSKKGK 344  
QY 350 VSEQLKCCSGILKEMFAKKAHAAYAWPFYKPDVVEALGLHDYCDIHKHPMDMTIKSKLEA 409  
DB 345 LSEQLKHCNGILKELLSKKAHAAYAWPFYKPDVVEALGLHDYCDIHKHPMDLSTVKRKME 404  
QY 410 REYRDAQEFAGADVRLMFSNCYKYNPPDHVWAMARKLDQVFMFAKMPDEPEEPVAVS 469  
DB 405 RYRDAQEFAGADVRLMFSNCYKYNPPDHVWAMARKLDQVFEFYAKMPDEPLEGPLV 464  
QY 470 SPAVPPP-TKVVAPPSSSDSSSS-----SDSDSTDDSEERAOHLAEQ 517  
DB 465 STALPGLTKSSSESSSESSSESSSESSSESSSESSSESSSESSSEERAHRLAEQ 524  
QY 518 LKAVHEQLAALSOPOONKPKKKEKKEKK-----EKHKKEVEENKKS-KAKEPP-PK 571  
DB 525 LRVAHEQLAALSOQPIKPKRK-REKKEKKKKAKEHGRIGIDEDDKGPRAPRPPQPK 583  
QY 572 KTKKNSSNSNV-----SKKEPAPMKSKPPPT-----YESEEDCKCKPMSE 613  
DB 584 KSKKAGGGSNATTLSPHGFSGSSNKLPPKSKQKAPPVLTGYDSEESRPMSE 643  
QY 614 EKRLSLDKLPGKGLGRVWHIIQSRPSLKNPNDETEIDFETLKPSTLRELGA-LCH 672  
DB 644 EKRLSLDKLPGKGLGRVWHIIQAREPSLRDSNPPEEIEIDFETLKPSTLRELYVLS 703  
QY 673 LLFAEEKETFKLRK 686  
DB 704 CLRKKPKPYTIRK 686



```

Db 256 EAKAATIPARRSGRPKPKKDLPSQQHQT-SKKGLSEQLKYNGIILKELLSKHAA 314
QY 372 YAMPFYKPDVEALGLHDYCDIIKHPMDMTIKSKLEAREYRDAQEFADVRLMFSNCYK 431
Db 315 YAMPFYKPDVASALGLHDYHEIIKHPMDLSTIKRKMENRDYHDAQEFADVRLMFSNCYK 374
QY 432 YNPDPDHEVAMARKLDQVFEMRFKAPKPDPEE-PVAVSSPAVPPPTKVVAPFSSSD--- 487
Db 375 YNPDPDHEVAMARKLDQVFESVAKMPDPEQDASPPSVSAPLIGALSSESSESSDEDD 434
QY 488 -----SSSDSSSDSDSDSEERAEORLAELQELKAVHEQLAALSQPOQNK 535
Db 435 EDEDEDEDEDESSSDSSSDSESS-DSEERANRLAEQLQELRAVHEQLAALSQGPVSK 493
QY 536 PKKKEKDKKKKKKKKEEVENKSKAKEPPPKTKK-----NNSNSNVNKKKEPAPM 591
Db 494 PKKREKKKKKKKKKKGRGDEE--SRARQAQLRKAAGGGGGGGSSKSKKAAGA 550
QY 592 KSRPPPT-YESEEDCKPKMSYEKKQLSLDINKLPCEKLGVRVHIIQSREPSLKNSNP 650
Db 551 ALPPPTLYDEEESESKPMTYDEKRLSLDINKLPCEKLGVRVHIIQSREPSLRDSNPE 610
QY 651 EIEDFETLKPSTLREL 667
Db 611 EIEDFETLKPSTLREL 627

RESULT 10
ID Q8QFT7 PRELIMINARY; PRT; 701 AA.
AC Q8QFT7;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DE Hypothetical 76.4 kDa protein.
GN BRD3
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21845882; PubMed=11856876;
RA Bouchireb N., Grutzner F., Haaf T., Stephens R.J., Elgar G.,
RA Green A.J., Clark M.S.;
RT "Comparative mapping of the human 9q34 region in Fugu rubripes.";
RL Cytogenet. Cell Genet. 94:173-179(2001).
DR EMBL; AJ311635; CAC84085.1;
KW Hypothetical protein.
SQ SEQUENCE 701 AA; 76356 MW; 74E1F328615E9BD0 CRC64;

Query Match 48.3%; Score 1757; DB 13; Length 701;
Best Local Similarity 56.4%; Pred. No. 7.1e-96;
Matches 372; Conservative 68; Mismatches 151; Indels 68; Gaps 15;

QY 26 MSTTQAAQPPAANAASTNPPPETSNPNKPKRTQLOQLVLLRVLTKLWKHFQFWFPQ 85
Db 1 MSDAPEAAPSP--PPLTNPPEVSNPNKPKRTQLOQLVLLRVLTKLWKHFQFWFPQ 58

QY 86 PVDVAVLNLDPDYKIIKTPMDMTIKRLENNYVNAQECIQDFNTFTNCYINPKGDD 145
Db 59 PVDAILCLADYHKVINKPNMDMTIKRLENNYVNAQECIQDFNTFTNCYINPKGDD 118

QY 146 IVLMAEALKFLQKINELPTEETEMIVQAKGRCRKETGTAKPCVSNPN---TTQA 202
Db 119 IVLMAQALEKIFLQKVAQMPQEEVALLPPAPKGNKSK-----QPAAGTYSQQAESSA 172

QY 203 STPPTQTPQPNPPVQATPHFPFVPTDILQTVPMVTVVPPVPPVPPVPPVPPVPP 262
Db 173 SSPPSYSPSPSTPWTIST-----TPTPVQTPPI-SAPQPPAA 210
```

```

QY 263 PAQPVOVSHPIIAATQPPVTKTKGVRKADTTPTT--IDPIHEPSPSPPEPKTKLG- 319
Db 211 MMP-----SAQPVVKKKGVRKADTTPTTSAISAGRADSPSAQDAKPAKLG 258
QY 320 -QRRESSRPVKPKP---KQVPSQOQHAPBPKSSKVSQKCCSGILKEMFAKHAAYAMP 375
Db 259 TRREATARPAKTRREVEEYAGGAGGGRKTKLGLQMKHCDAILKEMLSKHAAYAMP 318
QY 376 FYKPDVDEALGLHDYCDIIKHPMDMTIKSKLEAREYRDAQEFADVRLMFSNCYKNNP 435
Db 319 FYKPDAAEALGLHDYCDIIKHPMDLSTIRKMDKGEYNEQSPATDVRMFSNCYKNNP 378
QY 436 DHEVAMARKLDQVFEMRFKAPKPDPEEPVAVSSPAVPPPTKVVAPFSSSDSSSDSS 495
Db 379 DHEVAMARKLDQVFEMRFKAPKPDPEEPVAVSSPAVPPPTKVVAPFSSSDSSSDSS 432
QY 496 SDSTSDSEERAEORLAELQELKAVHEQLAALSQPOQNKPK-KKKEKDKKKKEK-KHR 553
Db 433 ESS-----DSEERATRLAEQLQELKAVHEQLAVLSQAPVSKPKKKKKKKKDKGNKA 489
QY 554 KEEVEENKSKAKEPPPKTKK--NNSNSNVNKKKEPAPMKSKPP---PTVESEEDCKK 608
Db 490 KMEEEKPKTKTAQPKPANQKPPARKANSTVTATROPKSSKTSVSGSTNGDDGESAL 549
QY 609 PMSYEKRLSLDINKLPCEKLGVRVHIIQSREPSLKNSNPDEIEIDFETLKPSTLREL 667
Db 550 PMSYDEKRLSLDINKLPCEKLGVRVHIIQTRPSLRDSNPDEIEIDFETLKPSTLREL 608

RESULT 11
ID Q8UUM2 PRELIMINARY; PRT; 814 AA.
AC Q8UUM2;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE RING3 protein.
GN RING3.
OS Oryzias latipes (Medaka fish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
OX NCBI_TaxID=8090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HD-RR;
RA Matsuo M.Y., Asakawa S., Shimizu N., Kimura H., Nonaka M.;
RT "Nucleotide Sequence of the MHC Class I Region of a Teleost, the
RL Medaka.";
RL Immunogenetics 0:0-0(2002).
DR EMBL; AB073376; BAB83842.1;
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR002052; N6_Mtase.
DR Pfam; PF00439; bromodomain; 2.
DR PRINTS; PR00503; bromodomain.
DR SMART; SM00297; BROMO; 2.
DR PROSITE; PS00633; BROMODOMAIN_1; UNKNOWN_2.
DR PROSITE; PS50014; BROMODOMAIN_2; 2.
DR PROSITE; PS00092; N6_MTASE; UNKNOWN_1.
SQ SEQUENCE 814 AA; 90116 MW; C71293789354623D CRC64;

Query Match 46.5%; Score 1690.5; DB 13; Length 814;
Best Local Similarity 49.6%; Pred. No. 6.9e-92;
Matches 377; Conservative 85; Mismatches 135; Indels 163; Gaps 21;

QY 1 MSAESGPGTBLRNLPMVMDGLTSSQMTTQAAQPPAANAASTNPPPETSNPNKPKRT 60
Db 24 MMDQSAAGKIRKPSLLYEDFESPGM-----PAISQMAPSGPPQPPVRDPSRHGRT 75

QY 61 NOQLYLLRVVLTKLWKHFQFWFPVPTDILQTVPMVTVVPPVPPVPPVPPVPPVPP 120
Db 119 IVLMAQALEKIFLQKVAQMPQEEVALLPPAPKGNKSK-----TPTPVQTPPI-SAPQPPAA 210
```

```
Db 76 NQLOFLQKVLKSLWRHFAWPEHPVDAAKLNLPDYHKLIKTPTMDMGTIRKLENNYR 135
Qy 121 NQAEICIDQFNFTNCYIYNKPGDDIVLMAEALFKLQKINELPTETETIMVOAKGRG 180
Db 136 SASCEQDMFNFTNCYIYNKPTDDIVLMAQSLKAFKQVAMPQEEIE--LPPPVPVS 193
Qy 161 RGRKETSATKQCVST--VPNTQTA-STPPTQTPQPPPPVQATPHFPFAVTPDLIVQT 236
Db 194 KOSKISG---GVTQAHQVAVSQSVSPP--TPE-----TPDSILST 230
Qy 237 PMVTI-----VPPQLOTPPPVPOPOPAPAPQVQSHPPPIIAATPQPVKTKGKVR 290
Db 231 PPQIILTKCSITLQPEQT--IPAITGPPPT-----QPTAKKKGKVR 270
Qy 291 KADTTPTTID-PI-----HEPP----- 307
Db 271 KADTTPTTAMPIMSTLGVGTIGLGMVGHDSPLLTSLGMDHSSSLGMMQAISISQGM 330
Qy 308 -----SLPEPKTKLQGRRESRPVKPKDVPDSQQHPAPEKSKVSEQLKCCSGI 360
Db 331 GGMGGGTMMLGAKATSGSRGVSGRPIKPKKOLPDSIL-PTPVRRSKLSPQLRYCSGV 389
Qy 361 LKEMFAKHAAYAMPFYKPDVVEALGLHDYCDIHKHPMDSTIKSLEAREYRDAQEFGA 420
Db 390 LKELLSKHAAYAMPFYKPDVAVSLGLHDYHEILKFPMDLSTIKRKMDCGREYRQAQFSA 449
Qy 421 DVRLMFSNCYKYNPPDHEVAMARKLDQVFEMFAKMPDEPEEPVAVSSPAVPPPTKV 480
Db 450 DVRLMFSNCYKYNPPDHDVAMARKLDQVFECFAKMPDE-----APAPSSST--- 497
Qy 481 APPSSSSSSSSSSSSDSTO-----DSEERARLAELQOLKAVHEQLAALSQ 530
Db 498 -----SSSSSSSSSELSUGSESSPSSDSEERANRLAELOQLKAVHEQLALSQ 553
Qy 531 PQONKPKK--EKDKERK--KKHKRKE-----VEENKSKAKEPKPKKT 573
Db 554 GPIVKPKKKKKKKKKRKEKRRHIEDLTPIRPKAPKTKTKTRKMDCPVLP 613
Qy 574 KKNSSNWNKKEPAPMKSP-----PYESEEEDCKPMSYEERKQLSLDINKLPG 627
Db 614 KKTOSKNSKSKSKAAITFNPHDPLVGHFDSDEEDTAPMSYDEKRLSLDINKLPG 673
Qy 628 EKLGRVHVHIIOSRPSLKNSPDEIDEIDFETLKPSTLREL 667
Db 674 EKLGRVHVHIIOSRPSLKDTPNEIDEIDEIDFETLKPSTLREL 713

RESULT 12
Q8T775 PRELIMINARY; PRT; 664 AA.
AC Q8T775;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 74.6 kDa protein (Fragment)
OS Branchiostoma floridae (Florida lancelet) (Amphioxus).
OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
OC Branchiostoma.
OX NCBI_TaxID=7739;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=11967531;
RA Abi-Rached L., Gilles A., Shlina T., Pontarotti P., Inoko H.;
RT "Evidence of en bloc duplication in vertebrate genomes."
RL Nat. Genet. 0:0-0(2002).
DR EMBL; AF391288; AAM18869.1; -.
KW Hypothetical protein.
FT NON_TER 664 664
SQ SEQUENCE 664 AA; 74627 MW; BF3912BA045253BF CRC64;

Query Match 46.28; Score 1680; DB 5; Length 664;
Best local similarity 50.9; Pred. No. 2,3e-91;
Matches 361; Conservative 87; Mismatches 133; Indels 128; Gaps 25;
```

```
Qy 1 MSASGPGTRLRNLPVM--GDGLETSMSTTQAQOPOPANAASTNPPPETSNPNKPKRQ 59
Db 1 MNSPTGAGKRRRPSIMYAEGETSE--SKQA-----PKKPGRM 37
Qy 60 TNQLOYLRLRVLKTWKHOFAPFPQOPVDVAKLNLPDYKKLIKTPTMDMGTIKRLENNY 119
Db 38 TNQLOYLKVVYKAVKHNFAWPEHPVDVYKLNIPDKIITKPTMDLGTIKRLETNY 97
Qy 120 WNAOECIQDNFTNCYIYNKPGDDIVLMAEALFKLQKINELPTETETIM-----IV 174
Db 98 YSAKECQDNFTNCYIYNKPGEDVYLMAGTLEKFLQKVAQMPPEVEVASQKPAV 157
Qy 175 QAKGRGRKGTAKPGCVSTVNTTQASTPPTQTPQPN-PPPVQATPHFPFAVTPDLI 233
Db 158 QTKGKKGPRKG-----PRVQAAS--TAVPVLNAQPPVQSPPPAPVVTPTPI 203
Qy 234 VQT-----PMTVVPQPL--QTTPPV-----PP-----QPQPPAPAPQPVQSH 271
Db 204 VPTTVAPTHVPTDLSLPQPIHQPPPVVQPPPTAPPARTQQPRPPPEQPPQLQMP 263
Qy 272 P-----PTIAATPO--PVKTKGKVRKADTTTPTTIDPIHEPPSLP--PEPKT 315
Db 264 PQTLPQGSTAPAPLISQPGQAVTKVKKGKRRADTTTPTTITTTGSESPSTTEPRP 323
Qy 316 TKLGO--RESSRPVKPKDVPDSQQHPAPEKSKVSQLCCSGILKEMFAKHAAYAW 374
Db 324 AKIPVRESGRQIKPPRRRLPETEQH--SSKKGKLSAQLKYCOGIKEMFAKHAAYAW 382
Qy 375 PFYKPDVVEALGLHDYCDIHKHPMDSTIKSLEAREYRDAQEFQADVRLMFSNCYKYN 434
Db 383 PFYEPDADLGLHDYHEILKHPMDLGTVKKKMDTREYKSAQEFASDMRMIFSNCRYNP 442
Qy 435 PDHEVAMARKLDQVFEMFAKMPDEPEEPVAVSSP---AVPPPTKVAP--PSSSDSS 490
Db 443 PEHDVQMARKLDQVFEMKAKMPDEPDP-----EPQPLVPPPPAEPENTSTSSSS 497
Qy 491 DSSSDSSDSDSDSEERARLAELQOLKAVHEQLAALSQ--PQONKPK--KKKEKDK---K 544
Db 498 DSSSDSDST--DSEERARLRELQDLQRLAVHEQLSALSQAQHKPKKKKKDKREK 555
Qy 545 EKKEKHKRKEEVENKSKAKE-----PPPKTKKKNSS 579
Db 556 DKDEKDKDKDKDKKHKKKKKKKKKEGEGVVVVKVDVAPTPVPPPPQPPKPKRQPK 615
Qy 580 NSNVSKKEPAPMKSK--PPPTYESEEDCKPMSYEERKQLSLDINKLP 626
Db 616 KTNSSRQSSKSKAQTTPAPAYESEDLDCKPMTYDEKRLSLDINKLP 664

RESULT 13
Q9GU61 PRELIMINARY; PRT; 732 AA.
AC Q9GU61;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE RING3 (Fragment).
GN RING3.
OS Myxine glutinosa (Atlantic hagfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Hyperotreti; Myxiniiformes;
OX Myxiniidae; Myxiniinae; Myxine.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUP-ACTIVATED PERIPHERAL BLOOD LEUKOCYTES;
RA White G.P., Cunningham C.;
RT "Characterization of RING3-like protein from Atlantic hagfish (Myxine glutinosa).";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF191032; AAG179.1; -.
DR HSP; 092831; 1891.
DR InterPro; IPR001487; Bromodomain.
```

DR Pfam: PF00439; bromodomain\_2.  
DR PRINTS: PR00503; bromodomain.  
DR SMART: SM00297; BROMO; 2.  
DR PROSITE: PS00633; BROMODOMAIN\_1; UNKNOWN\_1.  
DR PROSITE: PS00014; BROMODOMAIN\_2; 2.  
FT NON\_TER 732 732  
SQ SEQUENCE 732 AA; 81669 MW; 2038D61BD20F6B50 CRC64;  
  
Query Match 46.1%; Score 1678; DB 5; Length 732;  
Best Local Similarity 49.5%; Pred. No. 3,4e-91;  
Matches 376; Conservative 71; Mismatches 165; Indels 148; Gaps 19;  
  
QY 26 MSTTQAAQAPQANAASTNPPPTSPNPKPKROTNOLOYLRLVVLATLKKHQFAWPFQ 85  
Db 1 MATSPQOQQO---QPLCNPPPTPNPKPRLTNLOLQFLQKVMKVKHQFAWPFH 57  
  
QY 86 PVDVAVLNLDPYKIIKTPMDMTGKIKRLNNYVWNAQECIQDFTMTNCTYIYNKPGDD 145  
Db 58 PVDAAKLNLPDYQIKNPLDMLTIKRLSNYYTAVCEIQDFTMTNCTYIYNRND 117  
  
QY 146 IVLMAEALEKLFQKINELPTEETIMVQAK-CRGRGRKETGAKPGVSTVPTNTQAST 204  
Db 118 IVLMAOTVEKAFLOKVAEMPVEEYITSPVARVPQRRGRK-----PAALTAQA 165  
  
QY 205 PPQQTQPNPPVPOATPHPP-----AVTPDLIVQTPVMTVVPQPLQ----- 248  
Db 166 PVVTAGVPVLSPIPAAPSQRPRRAAAAATQOIAAATLPASRLRLBOCQGPVONQM 225  
  
QY 249 -----TPPPV--PPQPPAPAPQPVQSHPIIAATP----- 279  
Db 226 ARSVSTPPAVGQPTLETPSAAGVQSLLSQPMVSLTPLNMQGFAATHPTPKIDALGR 285  
  
QY 280 -----QPVTKKGVKVKADTTTPT 298  
Db 286 AERQGEEDTAMEQSELAFKGGRSGBGINPDSDISWVHTQENQASRGVKKKADTTTPT 345  
  
QY 299 TIDIPEPPSLPPPKTKLG-----QRESSRPVKPKKQVDSQO--HPA 343  
Db 346 AL-----QARPNEETSPGLADGPAGIVRPLVPGRESSRPVKARKEVPDPALPV 399  
  
QY 344 PEKSSKYEQKCCGILKEMFAKHAAYAWPFYKPYDVDEALGHDYCDIHKHPMDMTSI 403  
Db 400 SKRVQMSDQLRHQOTILKEFTKHAAYAWPFYKAVDAFALGHDYHDIIKIPMDLTTI 459  
  
QY 404 KSKLEAREYDAQFAGADVRLMFSNCKYKPNPDHEVYVAMARKLDQVFMFAKMPDE--- 460  
Db 460 KEKFEREYTNLHEFADDMRLMFSNCKYKPNPDHEVYVAMARKLDQVFMFAKMPDETPP 519  
  
QY 461 -PEPVAVVSSPAVPTTKVYAPP-----SSSDSSSDSSSDSDSDSDSEERAOELQ 515  
Db 520 PPVQPPVTPQPTTTPP-----PPDVHLTSSSSSSSDSDSDSEERAOELQ 574  
  
QY 516 EQLKAVHEQALQALQSQOONPKKKEKDKKKKK--HKRKEEVEENKSKAKPPPKKT 573  
Db 575 EQLKALHEQLASLSQAIPSKKK--KERKEKDKVLEKARKWEDEKVLKNLLQAKM 633  
  
QY 574 KKNSSNSNYSKE-----PAPKSKP--PPYSESEEDCKPMSYEKKRQLSLINKL 625  
Db 634 GKQNSNRPIKKEBLSDTIGASSLNAIPNVTTLPGDTEG-ARPMYDEKRLSLDINRL 692  
  
QY 626 PGEKLGVRVHIIQSRPSLKNPNDEIDFETLKPSTLR 665  
Db 693 PGDKLGVRVHIIQSRPSLKNPNDEIDFETLKPSTLR 732  
  
RESULT 14  
Q91Y44 PRELIMINARY; PRT; 956 AA.  
AC Q91Y44;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Bromodomain-containing female sterile homeotic-like protein.

GN FSRG3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=TESTIS;  
RA Shang E., Wolgemuth D.J.;  
RT "The Murine Fsr3 (Female Sterile Homeotic-Related Gene) Family  
RT Consists of at Least Four Members Whose Restricted Patterns of  
RT Expression Suggest Functions During Spermatogenesis";  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF358660; AAK50736.1; -;  
DR MGD; MGI:1891374; Fsr3.  
DR InterPro; IPR001487; Bromodomain.  
DR Pfam; PF00439; bromodomain; 2.  
DR PROSITE; PS00633; BROMODOMAIN\_1; UNKNOWN\_2.  
DR PROSITE; PS00014; BROMODOMAIN\_2; 2.  
SQ SEQUENCE 956 AA; 107338 MW; 8474CAEDC6217AC6 CRC64;  
  
Query Match 42.3%; Score 1540; DB 11; Length 956;  
Best Local Similarity 51.8%; Pred. No. 5,9e-83;  
Matches 332; Conservative 71; Mismatches 138; Indels 100; Gaps 14;  
  
QY 41 ASTNPPPTSPNPKPKROTNOLOYLRLVVLATLKKHQFAWPFQPVDAVKLNLPDYKI 100  
Db 9 AIVNPPPEYINTKSGRLTNQLQFLQRLVVLKALWKHGSWPFQPVDAVKLNLPDYTI 68  
  
QY 101 IKTPMDMGTIKRLNNYVWNAQECIQDFTMTNCTYIYNKPGDDIVLMAEALEKLFQK 160  
Db 69 IKTPMDLNTIKRLNNYVWNAQECIQDFTMTNCTYIYNKPGDDIVVMAQALEKLFQK 128  
  
QY 161 INELPTEETIMVQAKGRGRKETGAKPGVSTVPTNTQASTPPTQTPQPNPPVQA 220  
Db 129 LSQMPQEEVQV-----GKERIKDI-QOKIATVSSAKEQI-----PSKA 165  
  
QY 221 TPHPF-----PAVTPDLIVQTPVMTVVPQPLQTPPVFPQPPAPAPQVQSHPII 275  
Db 167 AENFKRQEPISGLPDISL-----SPLNMAQEAPP-----I 197  
  
QY 276 AATPQPVKTKGVKVKADTTTPTI---DPHEPPSPPEPKTKLGRRESSRPVKPK 332  
Db 198 CDSQSLVQITKGVKRRADTTTPTIXIAKASESPPTL-----RETQPNMPV 244  
  
QY 333 KD-----VPDSQHPAPEKSSKYSEQLKCCGILKEMFAKHAAYAWPFYKPYDVDEALG 386  
Db 245 KENTVKNVLPDSQGHKVLTKVTEQLKHCSEILKEMAKKHLPYAWPFYVNDADALG 304  
  
QY 387 LHDYCDIHKHPMDMTSKLEAREYDAQFAGADVRLMFSNCKYKPNPDHEVYVAMARKL 446  
Db 305 LHNVDYVKNPMDLGTIKGMDNOEYKDAVEFAADVRLMFMNCKYKPNPDHEVYVAMARKL 364  
  
QY 447 QDVFMFAKMPDEPEPVAVVSSPAVPTTKVYAPPSSSDSSSDSDSDSDSDSDSEEE 506  
Db 365 QDVFMFAKMPDEPEPVAVVSSPAVPTTKVYAPPSSSDSSSDSDSDSDSDSDSEEE 417  
  
QY 507 RAQRLAEQLQKAVHEQLAALQSQOONPKKKEKDKKKKK--HKRKEEVEENKSKAK 566  
Db 418 RVHLAKLQQLNVAHQLOQLVSOVPLRLKK--KNEKSRAPKRVKNNRDEPRKK-- 473  
  
QY 567 EPPPKTKKNSSNSNYSKEKPEAPMKSKPPPTTSEEDCKPMSYEKKRQLSLINKLP 626  
Db 474 ---PKMKQKEKAKINQPKK-----KPLKSEEDNAKPMYDEKRLSLINKLP 522  
  
QY 627 GEKLGVRVHIIQSRPSLKNPNDEIDFETLKPSTLR 667  
Db 523 GDLGVRVHIIQSRPSLKNPNDEIDFETLKPSTLR 563  
  
RESULT 15  
O14789 PRELIMINARY; PRT; 947 AA.  
ID O14789



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: March 1, 2003, 10:57:44 ; Search time 6007 Seconds

(without alignments)  
11283.575 Million cell updates/sec

Title: US-09-700-590A-101

Perfect score: 2329  
Sequence: 1 gcaaggatactacatgctc.....ngccttaaaagtlnccca 2329Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapevt 1.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

## Database :

GeneB1:.\*  
1: gb\_da.\*  
2: gb\_hlg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vl.\*  
15: em\_da.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vl.\*  
30: em\_hlg\_hum.\*  
31: em\_hlg\_inv.\*  
32: em\_hlg\_other.\*  
33: em\_hlg\_mus.\*  
34: em\_hlg\_pln.\*  
35: em\_hlg\_rod.\*  
36: em\_hlg\_mam.\*  
37: em\_hlg\_vrt.\*  
38: em\_sy.\*  
39: em\_higo\_hum.\*  
40: em\_higo\_mus.\*  
41: em\_higo\_other.\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2065	88.7	3149	9 HSHUNKI	Y12059 H. sapiens H
2	2065	88.7	3149	9 AF386649	AF386649 Homo sapi
3	1724	74.0	2727	10 AF461396	AF461396 Mus muscu
4	1722.4	74.0	6000	10 AF461395	AF461395 Mus muscu
5	1720.8	73.9	5281	10 AF273217	AF273217 Mus muscu
6	1256.2	53.9	2199	10 BC008532	BC008532 Mus muscu
7	663.2	28.5	3028	9 HUMOREX	D26362 Human mRNA
8	644.4	27.7	3033	10 BC031536	BC031536 Mus muscu
9	642.8	27.6	2933	10 AF269193	AF269193 Mus muscu
10	559.4	24.0	667	6 AX427508	AX427508 Sequence
11	541.6	23.3	2548	9 BC032124	BC032124 Homo sapi
12	461	19.8	3649	10 AF045462	AF045462 Mus sapi
13	460.2	19.8	4053	9 HSF5MR	X62083 H. sapiens m
14	460.2	19.8	4053	9 HUM5SHG	M80613 Human homol
15	460.2	19.8	4664	6 AX409010	AX409010 Sequence
16	460.2	19.8	4664	9 HUMKIA9001	D42040 Human mRNA
17	403.2	17.3	1554	10 AB010248	AB010248 Mus muscu
18	403.2	17.3	1657	10 AB010247	AB010247 Mus muscu
19	403.2	17.3	1688	10 AB010246	AB010246 Mus muscu
20	398.4	17.1	2206	10 AF318183	AF318183 Mus muscu
21	393.4	16.9	3391	10 AF358660	AF358660 Mus muscu
22	383.4	16.5	3429	9 BC011898	BC011898 Homo sapi
23	381.6	16.4	2350	9 BC000477	BC000477 Homo sapi
24	371	15.9	1884	9 AK024312	AK024312 Homo sapi
25	367	15.7	42497	9 AC004798	AC004798 Homo sapi
26	366.2	15.7	4667	9 HSM804033	AL837222 Homo sapi
27	336.6	14.5	2532	5 AF191032	AF191032 Myxine gl
28	311.2	13.4	177465	2 AC090975	AC090975 Mus muscu
29	311.2	13.4	235218	2 AC074208	AC074208 Mus muscu
30	298	12.8	40649	9 AC003111	AC003111 Human DNA
31	294.2	12.6	3106	9 AF019085	AF019085 Human sapi
32	275.6	11.8	207585	2 AC073779	AC073779 Mus muscu
33	255.6	11.0	236659	2 AL808021	AL808021 Mus muscu
34	253.6	10.9	3404	9 AK056504	AK056504 Homo sapi
35	208.2	8.9	1741	9 S78771	S78771 NAT-Cpg 1st
36	207	8.9	352	6 AX071305	AX071305 Sequence
37	203.8	8.8	5632	3 DROFSHA	M23221 D. melanogas
38	203.8	8.8	7792	3 DROFSHA	M23221 D. melanogas
39	203.8	8.8	37518	2 AC014003	AC014003 Drosophila
40	203.8	8.8	181178	3 AC023713	AC023713 Drosophila
41	203.8	8.8	306267	3 AE003442	AE003442 Drosophila
42	197.6	8.5	384	10 MMPJ2PFSH	X97573 M. musculus
43	170.2	7.3	4651	5 GGRING3GE	X96669 G. gallus RI
44	170.2	7.3	37644	6 A95302	A95302 Sequence 33
45	170.2	7.3	92863	12 GGBLOCUS	AL023516 Gallus ga

## ALIGNMENTS

RESULT 1  
LOCUS HSHUNKI 3149 bp mRNA linear PRI 02-MAY-1998  
DEFINITION H.sapiens HUNKI mRNA.  
ACCESSION Y12059  
VERSION Y12059.1 GI:3115203  
KEYWORDS HUNKI gene.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE  
AUTHORS Weber, B.  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 3149)

Pred. No. is the number of results predicted by chance to have a



AUTHORS Weber, B.  
TITLE Direct Submission  
JOURNAL Submitted (24-MAR-1997) B. Weber, Labor Paediatriche  
Molekularbiologie, Universitaetsklinikum Charite, Ziegelstr. 5-9,  
10098 Berlin, FRG

FEATURES  
source Location/Qualifiers

1.3149  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="placenta"  
223..2391  
/gene="HUNK1"  
223..2391  
/note="strong homology to human RING3 sequence"  
/protein\_id="CAA72780.1"  
/db\_xref="GI:3115204"  
/db\_xref="SPTREMBL:O60885"  
/translation="MSAESPGTRLDRLNLPVNGDLEISOMSTTQAOPOPAANA  
STNPPPTSPNPKPKRTNLOYLRLVLTLLKHOFPAMFQPOVDVAVKLNLDYKLIK  
TPDMGTIKRLLENNYWNAAECIDEDFTMTNCTYINPKPDIDVLALELEKLEIK  
INELPTEETLEIMVOAKGRGRKRETKAGKGVSTPNTQASTPOTOTPOPNPVP  
OATPPPEAVMPDLIVORPVMTVVPPOPIOTPPVPPOPOPAPAPAPPOVSHPIIA  
ATPOPVKTKKVKRKADPTPTTIDPIHEPSTLPEPKTKTGORRESRPYKPKKD  
VPDSQHPAPKSKSVKSVKPOLKCCSILKEMFRKHAATAMPYIKFVDEALGLHDYCD  
LIKHPMDSTIKSLKLEAREYDAQEFADVLMSKCYKNPDPHEVVAAMRKLDVFE  
EMRPAKDEDEEPEVAVSPVPPTVVAAPSSSDSSSDSSSDSSDSTDSSEERA  
ORLAIEQLAEVHQLAALSQPOONKKEKEDKKEKKEKHKKEVEENKSKAK  
EPPEKTKKNSNSNSNVSKEPKPKSKPPPYSEEDDKLPMYEERKROLSDINK  
LPKELGRVHVLTIOSRPSLTKNSPNDEIDEPITKLEPTEKLEHYVTSCLRRKKRPQ  
AEKVDYINGSSKMKMGFSSESSSESSSSSESSSESSSETGPA"

BASE COUNT 818 a 946 c 788 g 597 t  
ORIGIN

Query Match 88.7%; Score 2065; DB 9; Length 3149;  
Best Local Similarity 99.2%; Pred. No. 0; Mismatches 13; Indels 4; Gaps 3;  
Matches 2105; Conservative 0;

QY 4 GGGATCCTAGCATGCTCTCGGAGAGCGGCGCTGGAGAGATGGAATCTGCCAGTA 63  
DB 211 GGGATCCTAGCATGCTCTCGGAGAGCGGCGCTGGAGAGATGGAATCTGCCAGTA 270  
QY 64 ATGGGGGATGAGCTAGAAAATCTCCCAATGTCTACAAACAGGCCGCCAACCAG 123  
DB 271 ATGGGGGATGAGCTAGAAAATCTCCCAATGTCTACAAACAGGCCGCCAACCAG 330  
QY 124 CCAGCCAAAGCAGCAGCAACACCCCGCGCGCCGAGAGACCTCCCAACCTTAACAAGCC 183  
DB 331 CCAGCCAAAGCAGCAGCAACACCCCGCGCGCCGAGAGACCTCCCAACCTTAACAAGCC 390  
QY 184 AAGAGGCAAGCAGCAACCTGCAATACCTGTCAGAGTGCTGCTCAAGACATATGGAAA 243  
DB 391 AAGAGGCAAGCAGCAACCTGCAATACCTGTCAGAGTGCTGCTCAAGACATATGGAAA 450  
QY 244 CACCAAGTTGCGATGGCTTTCAGCAGAGCTGTGATGCCGTCAAGCTGACCTCTGAT 303  
DB 451 CACCAAGTTGCGATGGCTTTCAGCAGAGCTGTGATGCCGTCAAGCTGACCTCTGAT 510  
QY 304 TACTTAAGATCAATTAAGCGCTATGATGGGAACAATAAGAAAGCGCTGGAAAC 363  
DB 511 TACTTAAGATCAATTAAGCGCTATGATGGGAACAATAAGAAAGCGCTGGAAAC 570  
QY 364 AACTATTACTGGAATGCTCAGGAATGATCAAGACTTCAACACTATGTTTCAAAATGT 423  
DB 571 AACTATTACTGGAATGCTCAGGAATGATCAAGACTTCAACACTATGTTTCAAAATGT 630  
QY 424 TACATCTACAAACAGCTGAGATGACATAGTCTTAATGGCAGAAAGCTTGGAAAAGCTC 483  
DB 631 TACATCTACAAACAGCTGAGATGACATAGTCTTAATGGCAGAAAGCTTGGAAAAGCTC 690  
QY 484 TTCTTGCAAAAAATAATAGCTACCCACAGAAAGAAACCGAATCATGTAGTCCAGGCA 543

DB 691 TTCTTGCAAAAAATAATAGCTACCCACAGAAAGAAACCGAATCATGTAGTCCAGGCA 750  
QY 544 AAGAGAAAGAGAGCTGGAGAGAAAGAAACAGAGAGCAAAACCTGGCTTTCACAGGTA 603  
DB 751 AAGAGAAAGAGAGCTGGAGAGAAAGAAACAGAGAGCAAAACCTGGCTTTCACAGGTA 810  
QY 604 CCAAAACACACTCAAGCATGACTCTCTCCGAGAGACCCAGACCCCTAGCCGAATCTCTCT 663  
DB 811 CCAAAACACACTCAAGCATGACTCTCTCCGAGAGACCCAGACCCCTAGCCGAATCTCTCT 870  
QY 664 CCTGTGAGAGGACGAGCTACCCCTTCCCTGCGTACACCCGGAGCTCATGTGCGAGAC 723  
DB 871 CCTGTGAGAGGACGAGCTACCCCTTCCCTGCGTACACCCGGAGCTCATGTGCGAGAC 930  
QY 724 CCTGTATGACAGTGTGCTCTCCAGCAGCTGACAGAGCCCGCCAGTGGCCCCCAG 783  
DB 931 CCTGTATGACAGTGTGCTCTCCAGCAGCTGACAGAGCCCGCCAGTGGCCCCCAG 990  
QY 784 CCACAACCCCAACCCGCTCCAGCTCCCAAGCCGCTACAGAGCCACCCATCATTCGCG 843  
DB 991 CCACAACCCCAACCCGCTCCAGCTCCCAAGCCGCTACAGAGCCACCCATCATTCGCG 1050  
QY 844 GCACCCCAACAGCTGTGAAAGCAAAAGAGAGGAGGAAAGGAGAGACACCAACACC 903  
DB 1051 GCACCCCAACAGCTGTGAAAGCAAAAGAGAGGAGGAGGAAAGGAGAGACACCAACACC 1110  
QY 904 CCACACCATTTGACCCCATTCAGAGAGCACCCTGCTCCCGGAGCCCAAGACACC 963  
DB 1111 CCACACCATTTGACCCCATTCAGAGAGCACCCTGCTCCCGGAGCCCAAGACACC 1170  
QY 964 AAGCTGGGCGACCGCGGAGAGAGCAGCCGCTGTGAACCTTCAAAGAGAGCTGCC 1023  
DB 1171 AAGCTGGGCGACCGCGGAGAGAGCAGCCGCTGTGAACCTTCAAAGAGAGCTGCC 1230  
QY 1024 GACTCTCAGCAGCAGCCAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1083  
DB 1231 GACTCTCAGCAGCAGCCAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1290  
QY 1084 TGCAGCGGATCCTCAAGAGATGTTTTCGCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1143  
DB 1291 TGCAGCGGATCCTCAAGAGATGTTTTCGCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1350  
QY 1144 TCAAGAGCTGTGAGAGCTGAGAGCACTGGGCTTACACAGACTGTGACATCAAGAC 1203  
DB 1351 TCAAGAGCTGTGAGAGCTGAGAGCACTGGGCTTACACAGACTGTGACATCAAGAC 1410  
QY 1204 CCCATGAGATGAGCAGCAATCAAGTCTAACTGGAGGCGCCGATACCGTATGCTGAG 1263  
DB 1411 CCCATGAGATGAGCAGCAATCAAGTCTAACTGGAGGCGCCGATACCGTATGCTGAG 1470  
QY 1264 GAGTTGTGCTGAGCGTCGATGATGTTCTTCAACTGCTATAGTACAACTCTCTGAC 1323  
DB 1471 GAGTTGTGCTGAGCGTCGATGATGTTCTTCAACTGCTATAGTACAACTCTCTGAC 1530  
QY 1324 CATGAGGTGTGAGCGATGAGCGCCGCAAGCTCCAGAGATGTGTTGAATGCGCTTTCAG 1383  
DB 1531 CATGAGGTGTGAGCGATGAGCGCCGCAAGCTCCAGAGATGTGTTGAATGCGCTTTCAG 1590  
QY 1384 ATGCCGAGAGAGCTGAGAGAGCAGAGGTGAGCGGTCTCTCCCGAGTGGCCCTCC 1443  
DB 1591 ATGCCGAGAGAGCTGAGAGAGCAGAGGTGAGCGGTCTCTCCCGAGTGGCCCTCC 1650  
QY 1444 ACCAAGTTGTGGCCCGCTCATCTACAGCAGCAGCAGCAGAGATAGTCTCTCGAGAGT 1503  
DB 1651 ACCAAGTTGTGTGGCCCGCTCATCTACAGCAGCAGCAGCAGAGATAGTCTCTCGAGAGT 1710  
QY 1504 GACAGTTGAGTGTGACTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1563  
DB 1711 GACAGTTGAGTGTGACTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1770  
QY 1564 CAGCTCAAAAGCGGTGACAGAGAGAGCTGTGAGCCCTCTCAGCCCGCAGAGAAACCA 1623  
DB 1771 CAGCTCAAAAGCGGTGACAGAGAGAGCTGTGAGCCCTCTCAGCCCGCAGAGAAACCA 1830









QY	1861	CGGCGCTCAGCTTGGCAATCAACAAGCTCCCGGCGAGAACCTGGGCGCGTGGTGCAC	1920
Db	2090	CGGCGCTTAAGTCTAGATATCAACAACCACTCCTGGTGGAGAACTAGCGCGTTAGTACAC	2149
QY	1921	ATCATCCAGTCAAGGAGCCCTCCCTGANAATTTCCAAACCCGACGAGATTGAATCGAC	1980
Db	2150	ATAATTCAGTCAAGGGAACCATCACTTAAAACTCCAAACCCGAGAGATTGAGATTGAC	2209
QY	1981	TTTGGACCCCTGAACCCGTCACACTTGGGTGAGCTTGGAGCGCTATGTCACCTCCGTGTT	2040
Db	2210	TTTGGACCCCTGAACCCATCTACACTACGAGAG - TTGAGCGATATGTCACTCCTCGTGT	2268
QY	2041	GCGGAGAAAAGAAAACCTTCAGCTGAGAAAGTTGATGTGATWTTGCCGGGTTCTCCAA	2100
Db	2269	GCGGAGAAAAGAAAAC - TCAGCTGAAAAAGTTGACGTGAT - -GCTGGTTCTCCAA	2325
QY	2101	NATGAAGGNTTCTCGGTCTCAAG 2125	
Db	2326	GATGAGGGATTCATCCTCTGAG 2350	

LOCUS	6000 bp	mRNA	linear	ROD 13-MAY-2002
DEFINITION	AF461395			
REFERENCE	Mus musculus bromodomain-containing protein BRD4 long variant (Brd4) mRNA, complete cds.			
VERSION	AF461395			
KEYWORDS	AF461395.1	GI:18308124		
SOURCE				
ORGANISM	Mus musculus.			
REFERENCE	Mus musculus.			
ATTNORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 6000)			
TITLE	Houzelstein, D., Bullock, S.L., Lynch, D.E., Grigorleva, E.F., Wilson, V.A. and Beddington, R.S. Growth and early postimplantation defects in mice deficient for the bromodomain-containing protein Brd4			
JOURNAL	Mol. Cell. Biol. 22 (11), 3794-3802 (2002)			
MEDLINE	21993140			
PUBMED	11997514			
REFERENCE	2 (bases 1 to 6000)			
AUTHORS	Houzelstein, D.			
TITLE	Direct Submission			
JOURNAL	Submitted (19-DEC-2001) Genetique et Developpement, Institut Jacques Monod, 2 Place Jussieu, Paris 75 251, France			
FEATURES				
source	Location/Qualifiers			
	1..6000			
	/organism="Mus musculus"			
	/db_xref="taxon:10090"			
	/chromosome="17"			
	/map="between D17Mit181 and D17Mit191"			
	1..6000			
gene	/gene="Brd4"			
	242..4444			
CDS	/gene="Brd4"			
	/codon_start=1			
	/product="bromodomain-containing protein BRD4 long variant"			
	/protein_id="AA167833.1"			
	/db_xref="GI:18308125"			
	/translation="MSTESGSGFRLRLPVWGDLFETSONSTGQAQOPAPANAATNTNPPEPTSNPKPKKQTNQLOYLIVLVEKTLTKHQFPAPQOPAVANKLIDYKILIKKIPDMGTIKKRLENNYVMAOECIDQENMETNTNCYINKRGDDIVLMAELKFLQKINLEPTEELIMTVQAGRGGRKRETAARAGSVSTVNTQASISPOTFQOQPPPPVATTFHPFPAVTPDLIAQPFVMTVNPQPLDTBSVPQPPPPAVPVQVSHPIIILIAATTPQVTKRGVKRAKDTTPTTIDIEHPESLAPBPKTAKLGRRESSRPVPPKPKDVSQDHPPEKSKISDQLCCSGILKEFAKHAAYAMPYQVDEVLGLHDYCDIILKHPDMSTIKSLKESREYDQAEAGDVRLEMSKYKNPDHVEVMAKRLQDVFEMRFAKMDPEEPYVTVSSPAVPTPTKYVAAPSSSDSSDSSDSSDSSDSEERAGRLAEQEDLKAVHEDLALSPQQKPKKKEKKKKKKKKKKKKKEEYENKSKSTKELPPKTKTKNNSSNSVSKPEVPTTKQPPYSESEEDKKKMSYEENKSLDINT			

BASE COUNT  
ORIGIN

1624 a 1835 c 1380 g 1161 t

	Query Match	74.0%;	Score 1722.4;	DB 10;	Length 6000;
	Best Local Similarity	89.4%;	Pred. No. 0;		
	Matches 1899;	Conservative	0;	Mismatches 219;	Indels 7; Gaps 4;
OY	4	GGAATCACTAGCATGTGCTGGCGAAGCGGCCCTGGGACGAGATTGAGAATTCGTCCAGTA	63		
Db	230	GGGATCTACTAGCATGTCTACGAGAGCGGCCCTGGGACAAATTGAGAAATTCGCACATA	289		
OY	64	ATGGGGGATGGACATGAAGAACCTTCCCAAATGTCTCAACAACAGGGCCGACGCCAACCCGAG	123		
Db	290	ATGGGGGATGGACATGAAGAACCTTCCCAAATGTCTCAACAACAGGGCCGACGCCAACCCGAG	349		
OY	124	CCAGCCCAACGCGAGCAGCACCMACCCCCCGCCCGACAGACCTCCMACCTTAACAAAGCC	183		
Db	350	CCAGCAATATGAGCAGCAGACCACCAATCTCTCCACCCCAGAGACTCCAAACCTTAACAAAGCC	409		
OY	184	AAGAGCAGACCAACCAACTGCAATPACCTGCTCAGAGTGTTGCTCAGACACTATGAAA	243		
Db	410	AAGAGCAGACCAACCAACTGCAATATCTGCTCAGAGTGTTGCTCAGACACTATGAAA	469		
OY	244	CACAGATTGATGAGCCTTTCACAGCGCTTGATGGCGTCAAGCTCAACCTCCAT	303		
Db	470	CACAGATTGATGAGCCTTTCACAGCGCTTGATGGCGTCAAGCTCAACCTCCAT	-529		
OY	304	TACTATAGATCATTAATAAACGCTATGATATGGGAACAAATAAGMAAGCGTTGGAAAA	363		
Db	530	TACTATAGATCATTAATAAACGCTATGATATGGGAACAAATAAGMAAGCGTTGGAAAA	589		
OY	364	AACATATTACGGGATGCTCAGAGATGTATCCAGGACTTCAACACTATGTTTACAAATGT	423		
Db	590	AACATATTACGGGATGCTCAGAGATGTATCCAGGACTTCAACACTATGTTTACAAATGT	649		
OY	424	TACATCTACAAACAGCCTGAGATGACATATCTTAATGGCAGAAAGCTCTGSAANAAGTC	483		
Db	650	TACATCTACAAACAGCCTGAGATGACATCTCTTAATGGCAGAAAGCTCTGSGAAGCTC	709		
OY	484	TTCTTGCAAAAAATTAATGAGCTACCCACAGAGAAGAAACCGAGATCATGATAGTCCAGCA	543		
Db	710	TTCTTGCAAAAAATTAATGAGCTACCTCTACAGAAAGAAACTGAGATCATGATAGTCCAGCA	769		
OY	544	AAAGSAGAGSAGCCTGGGAGAGAAAGAAACAGSAGACAGCAAACTGGCGTTCCAGGTA	603		
Db	770	AAAGSAGAGSAGCAGGAGAGAAAGAAACAGGAGGACAGAAAGCCTGGGTATCCAAGTA	829		
OY	604	CCAACCAACAACTCAAGATGAGCTCTCCGAGACCCAGACCCCTCAAGCCAA--TGCT	660		
Db	830	CCAACCAACAACTCAAGATCAACCTTCTCCGAGACCCAGACCCCTCAAGACCCCTCTCT	889		
OY	661	CCCTCTGTGAGGCGCAGCCTCACCCCTTCCCTGCGCTCACCCCGAGACTCATGCTCAG	720		
Db	890	CCACCTGTGAGGCGCACAACCTCACCCCTTCTCTCTGTCAACCCAGACCTCATGTGCCAG	949		
OY	721	ACCCCTGTATGATAGATGATGTGCTCCCGACGACATGCAAGAGCGCCCGCAGTGCCTCC	780		
Db	950	CCCTCTGTATGATAGATGATGTGCTCCCGCAGCCTTCAAGATCTCTTCAACCGGTACCTCC	1009		







Qy	1921	ATCATCCAGTACACGGAGACCCCTCCTCCCTGAAAGATTTCACACCCCGACAGATTTGAATGCAC	1980b
Db	2089	ATTAATTCAGTCAAGGAAACCATCACTCTTAAAAACCCCGATGAGATTGTAC	2148
Qy	1981	TTTGAGACCCCTGAAGCCGCTCCACACTGGGTGAGCTTGGAGGCGATATGTCACCTCCTGTTT	2040
Db	2149	TTTGGAGACCCCTGAAACCCATCTACACTACGAGAGG-TTGGAGCGATATGTCACCTCCTGTTT	2207
Qy	2041	GCAGAGAAAAGAAACCTTCAGAGCTGAGAAAGTTGATGTGATNTGCCGGGTCTCTCCA	2100
Db	2208	GCAGAGAAAAGAAAGAAC-TCAGCTGAGAAAAGTTGACGTGATT--GCTGGTTCTTCCAA	2264
Qy	2101	NATGAAGGNTTCTCGGTCTCAGG	2125
Db	2265	GATGAAGGATTTCTATCTCTGAG	2289

RESULT 6	BC008532	LOCUS	DEFINITION	ACCESSION	VERSION
	BC008532	2199 bp	mRNA		
		Mus musculus, clone IMAGE:3489640, mRNA.			
	BC008532				
	BC008532.1	GI:14250219			

REFERENCE  
1 (bases 1 to 2199)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus

Submitted (25-MAY-2001) National Institutes of Health, Mammalian

REMARK	COMMENT
NIH-MGC Project URL: <a href="http://mgc.ncl.nih.gov">http://mgc.ncl.nih.gov</a>	Contact: MGC help desk

Email: cgarbbs-r@mail.nih.gov  
Tissue Procurement: lothar Hennighausen Ph.D., Robin Humphreys  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www-shgc.stanford.edu>  
Contact: (Dickson, Mark) mcd@pdxil.stanford.edu  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAK Family: 7 Row: b Column: 12  
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, similarity but not identity to protein.

**FEATURES**  
**SOURCE**

```

/organism="Mus musculus"
/db_xref="taxon:10090"
/map="C57BL/6J"
/clone="IMAGE:3489640"
/tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months
old. gross tissue."
/clone_id="NCI_CGAP_Mam5"
/lab_host="DH10B"
/note="Vector: pCMV-Sport6"
BASE COUNT      611 a      642 c      542 g      404 t
ORIGIN

```

Query Match	53.9%	Score 1256.2	DB 10	Length 2199
Best Local Similarity	88.6%	Pred. No. 1.7e-264		
Matches 1373; Conservative	0	Mismatches 173;	Indels 3;	Gaps 1;

QY	1119	CCAGCCAGCCAAAGCAGCCAGCCAAACCCCCCCCCCCAGAGACTCCAAACCTTAACA	178
Db	647	CTGTGGCAGCCCTTGAGGATGTGGAAACCCACTAGGGGGGCTCCAGAGACTCCAAACCTTAACA	706
QY	1179	AGCCCAAGAGGCACACCAACCAACTGCAATAACCTGCTCAGAGAGTGGGTCTCAAGACACTAT	238
Db	707	AGCCCAAGAGGCACACCAACCAACTGCAATAATCTGCTCAGAGAGTGGGTCTCAAGACACTAT	766
QY	239	GGAAACACAGTGTTCATGTGCCTTTCAGCAGCGCTGTGATGTGCCTGCAAGCTGAACCTTC	298
Db	767	GGAAACACAGTGTTCATGTGCCTTTCAGCAGCGCTGTGATGTGCCTGCAAGCTGAACCTTC	826
QY	299	CTGATTTACTATAGATCATTTAAAAAGCCTTATGGATTTGGCAACATATAAAGACCGCTTGG	358
Db	827	CTGATTTACTATAGATCATTTAAAAAGCCTTATGGATTTGGCAACATATAAAGACCGCTTGG	886
QY	359	AAAACACTATTTACTGGAATGCTCAGAAATGTATCCAGGACTCCAACTATGTTTACAA	418
Db	887	AAAACACTATTTACTGGAATGCTCAGAAATGTATCCAGGACTCCAACTATGTTTACAA	946
QY	419	ATTGTACATCTTACAAACAGCCTGTGAGATGACATAGTCTTATATGGCAGAGAGCTGTGAAA	478
Db	947	ATTGTACATCTTATTAACAAAGCCTGTGAGATGACATGCTCTTAATGGCAGAGAGCTGTGAGA	1006
QY	479	AGCTCTTCTTGCACAAAATTAATATGAGCTACCCACAGACAGAAACCGATCATGATATGCC	538
Db	1007	AGCTCTTCTTGCACAAAATTAATATGAGCTCTACAGAAAGAACTGATCATGATATATGCC	1066
QY	539	AGGCAAAAGGAAGGAGCGTGGGGAGAAACAAACAGGACAGCAAAACCTGGCGTTTCCA	598
Db	1067	AGGCAAAAGGAAGGAGCGTGGGGAGAAACAAACAGGACAGCAAAACCTGGTGTATTTCCA	1126
QY	599	CGGTACCAAAACACAACTCAAGCATTCGACTCTTCGGAGACCGACAGACCCCTCAGCCGAAA	656
Db	1127	CGGTACCAAAACACAACTCAAGCATTCGACTCTTCGGAGACCGACAGACCCCTCAGCCGAAA	1186
QY	657	TCTCTCTCTGTGACAGGCCACGCGCTACCCCTTCCCTGTGCGCTACCCCGGACCTCATGCG	715
Db	1187	CTCTCTCTCTGTGACAGGCCACCAACTACCCCTTCTCTGTGACACCCGACCTCATGCTATG	1246
QY	716	TCGACAGCCCGTGATACAGTGGTGGCTCCCGAGCACTGCAGACGCCCCCGCCAGATGC	775
Db	1247	CCGAGCCTCTGTATACATAGTGGTGGCCCTCTAGCGACTCTCAGATCTCTTACCGGTATC	1306
QY	776	CCCCCAGCCACAAACCCCGCCAGCTTCAGACTCCCCAGCGCCGTACAGACCCACCCCAACA	835
Db	1307	CCCCCAGCCACAAACCCCGCCAGCTTCAGACTCCCCAGCGCCGTACAGACCCACCCCGCCA	1366
QY	836	TCATATCGCGCCACCCCGACAGCTGTGAAGCAAAAGAGGAGTGAAGAGGAAAAGCAGACA	895
Db	1367	TCATATCGAGCACCCCGCCAGCTGTGAAGCAAAAGAGGAGTGAAGAGGAAAAGCAGATA	1426
QY	896	CCACACACCCCAACACCATTTGACCCCATTTACAGAGCCACCTTCGCTGCCCCCGGAGGCCA	955
Db	1427	CCACACACCCCAACACCATTTGACCCCATTTATAGCCACCTTCACCTGCGGCCACAGAGCCCA	1486
QY	956	AGACACCAACAGTGTGGGCAACGGCGGGAGAGCAGCGGCGCTGTGAAAACCTTCCAAAGAAAG	1013
Db	1487	AGACACCAACAGTGTGGGCAACGGCGGGAGAGCAGAGACTGTGAAAACCTTCCAAAGAAAG	1546
QY	1016	ACGTGCCGCACTTCAGCAGCAGCACCACAGCAGAGAAAGACAGCAAGTCTGTGGAGCAGC	1075
Db	1547	ATGTACCGGAGCTACAGCAGCAGCACCAGGGCCAGAGAAAGACAGCAAGATCTCTGAGCAGC	1606
QY	1076	TCAGATGCTGTACAGGGCATCTCAAGAGATGTTTTCCAAAGAACAGCGCGGCTACGCGCT	1133
Db	1607	TAAAGTCTGTACAGGGCATCTCAAGAGATGTTTTCCAAAGAACAGTGTGCTTGTGCTT	1666
QY	1136	GGCCCTTCTTCAAGCCGTGTGGAGCTGTGGAGCACTTGGGCTTACACGCACTCTGTGCATATCA	1195
Db	1667	GGCCCTTCTTCAAGCCGTGTGGAGCTGTGGAGCACTGGGCTTGCACGCACTCTGTGCATATCA	1726
QY	1196	TCAAGCACCCCATGAGCATGAGACAAATCAAGTCTAACTGAGAGCCCGTGAATACGCTG	1255

Db	1727	TCAAATCATCCATGGACATGAGCACAATCAAGTCTAAACTAGATGCCGAGATACAGAG	1786
Qy	1256	ATGCCTCAGAGATTGGCTGCTGACTCCGATGATGTTCTCCAACTGCTATTAAGTACAAACC	1315
Db	1787	ATGCCACGAAATTTGGTCTCTATCTCCGATTGATGTTCTCAACTGCTACAACTAACACC	1846
Qy	1316	CTCTGACCATGAGTGTGGTGGCCATGAGCCCGCAAGCTCCAGATATGTTGCAATGGCT	1375
Db	1847	CCCCGACCATGAAGTGTGTAGCCATGGCTCGAAAACCTCCAGATATGTTGAAATCCGCT	1906
Qy	1376	TTGCCAAGATCCCGGACGAGCCTTAGAGACCACTGGTGGCCGTGCTCCCGCGCAGTGC	1435
Db	1907	TTGCCAAGATCCCGGAGAGCCTTAGAAGCCAGTTGTTAAAGTCTCTCCGCGCAGTGC	1966
Qy	1436	CCCCCTCCCAAGATTTGTGGCCCGCCCTCATCTACAGCAGACAGACAGATAGCTCCCT	1495
Db	1967	CACCCCTCAACAAAGTGTGTAGCCACCACCTCATCTATGATGACAGCAGCAGCAAGTTCTT	2026
Qy	1496	CGGACAGTGAAGTTCGACCTGATGATCTCTGAGAGAGCAGCCAGCCGCTGCTGAGC	1555
Db	2027	CCGACAGGAGACAGTTTCCACTGACCTCTGAGAGAGCAGCCAGCCGCTGCTGAGC	2086
Qy	1556	TCGAGGAGCAGCTCAAAAGCCGTGACAGAGCAGCTTGAGCCCTCTCAGCCCAACAA	1615
Db	2087	TCGAGGAGCAGCTCAAAAGCCGTGATGAGAGCAGCTTGAGCCCTCTCAGCCCAACAA	2146
Qy	1616	ACAACCAAGAAAAAGAGAGAAACAAGAAAGAAAAAGAAAAA	1664
Db	2147	ACAACCAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAA	2195
RESULT 7			
LOCUS	HUMORFX	3028 bp	linear PRI 06-OCT-2001
DEFINITION	Human mRNA for KIAA0043 gene, complete cds.		
ACCESSION	D26362		
VERSION	D26362.1	GI:452518	
KEYWORDS	KIAA0043		
SOURCE	Homo sapiens male myeloblast cell_line K5-1 cDNA to mRNA.		
ORGANISM	Homo sapiens		
REFERENCE	1		
AUTHORS	Nomura, N., Nagase, T., Miyajima, N., Sazuka, T., Tanaka, A., Sato, S., Seki, N., Kawarabayashi, Y., Ishikawa, K. and Tabata, S.		
TITLE	Prediction of the coding sequences of unidentified human genes. II. The coding sequences of 40 new genes (KIAA0041-KIAA0080) deduced by analysis of cDNA clones from human cell line K5-1		
JOURNAL	DNA Res. 1 (5), 223-229 (1994)		
MEDLINE	96051398		
REFERENCE	2 (bases 1 to 3028)		
AUTHORS	Ohara, O., Nagase, T., Kikuno, R. and Nomura, N.		
TITLE	Direct Submmission		
JOURNAL	Submitted (27-DEC-1993) Osamu Ohara, Kazusa DNA Research Institute; 1532-3, Yama, Kisarazu, Chiba 292-0812, Japan		
FEATURES	(E-mail: cdna@info.kazusa.or.jp, Tel: +81-438-52-3913)		
SOURCE	1. 3028		
gene	/organism="Homo sapiens"		
5'UTR	/db_xref="taxon:9606"		
CDS	/sex="male"		
	/cell_line="K5-1"		
	/cell_type="myeloblast"		
	1. 3028		
	/gene="KIAA0043"		
	1. 139		
	/gene="KIAA0043"		
	140. 2320		
	/note="similar to Human homolog of Drosophila female		
	sterile homeotic mRNA ( HUMSHG)"		
	/codon_start=1		
Qy	844	GCGACCCACAGCCTGTGAAGACAAAGAAAGGAGTAAAGAAAGACAGACACCAACG	903
Db	770	GTCACAGTCCCGCCAGCTGGCCGACCACTCTCTGCGACACCACTTCCTCCGCTGGTC	829
Qy	844	GCGACCCACAGCCTGTGAAGACAAAGAAAGGAGTAAAGAAAGACAGACACCAACG	903
Db	770	GTCACAGTCCCGCCAGCTGGCCGACCACTCTCTGCGACACCACTTCCTCCGCTGGTC	829



```

Db      830 CCTCTACGCGCGCTGTGTGAAGAAAAGGCGGAAAGCGAAGACGACACAACTCT 889
Qy      904 CCCACCAACATGACCCCATTCAC-----GAGCCACCCTGCTGCCCCGGAGACC 954
Db      890 CCCAGAGAGTGGGCGATCAGCGACGCGGAGTAGTGTGCCCCCGCTTGTACAGACCC 949
Qy      955 AAGACCAACAGTGGGCGACGCGGAGAGCA---GCGGCGCTGTGAACCTCCAAAG 1011
Db      950 AAGCAGGCGAAAGTGTGGCCCGGCGAGAGTGTGGCCCGCCATCAACCTCCCAAG 1009
Qy      1012 AAGCAGGCGCGAGCTGTCAAGCAGACCCAGACAGAAAGAGCAGCAAGTCTCGAG 1071
Db      1010 AAGCAGCTGGAGAGCGCGGAGGCGCCAGCAGCAGGAGAAAGAGGAACTGTGCGAG 1069
Qy      1072 CAGCTCAAGTGTCTCAGCGCGATCTCAAGAGAGATGTTGGCAAGAACACGCGCTAC 1131
Db      1070 CAGCTGGGCTACTGCGACAGCATCTCAAGAGAGATGCTATCAAGAAAGACAGCGGCTAC 1129
Qy      1132 GCGTGGCCCTTCTACAAAGCTGTGAGCGTGAAGCGACTGGGCTACACGACTGTGTAC 1191
Db      1130 GCGTGGCCCTTCTACAAAGCTGTGAGCGTGAAGCGCGCTGTGAGCTACACAGAC 1189
Qy      1192 ATCATCAAGCAACCCCATGAGATGATGACACATCAAGTCTAACTGGAGGCCGCTGATAC 1251
Db      1190 ATCATCAAGCAACCCCATGAGATGATGACACATCAAGTCTAAAGATGAGGCGGAGATAC 1249
Qy      1252 CGTATGCTCAGAGATGTTGGTGTGACGTGCGATGATGTTCTCAACTGCTATTAAGTAC 1311
Db      1250 CCACACGACAGGCGCTTGTGCTGTGATGTCCGCTGATGTCTGAAATGCTTACAAATAC 1309
Qy      1312 AACCTCTCTGACATGAGGTGTGCGCATGAGCGGACGCTCAAGATGTGTGAAATG 1371
Db      1310 AATCCCCAGACACGAGGTGTGGCCATGCGCCGAAAGCTCAAGAGCTGTGTGAGATG 1369
Qy      1372 CGTGTCCAAAGATGCGCGAGAGCGCTGAGAGCGCAGTGTGGCGCTGTCTCCCGGCA 1431
Db      1370 AGGTTGCGAGATGCGAGATGAGCGCGGTGAGGACCGG-----GCTGTGCT 1417
Qy      1432 GTGCCCCCTCCACCAAGATGTTGTGCGCCCGCTCATCAGCAGCAGCAGCAGGATAGC 1491
Db      1418 GCGCCCGGCGCGCGCATGTGTGAGCAGAGGCGCTGAGAGCAGCCGTAGCAGTGAAGAGC 1477
Qy      1492 TCTCTGAGACATGACATGCTGATGACTCTGAGAGAGCAGCGCCAGCGCTGTGCT 1551
Db      1478 TCTCTGAGACTGAGCAGCTGTG-----GACTCGGAGAGAGCGCGCCACAGCGCTGTGCG 1531
Qy      1552 GAGCTCGAGAGCAGCTCAAAAGCGCTGACAGCAGCAGCTTGCAGCCCTCTCTCAGCCCGAG 1611
Db      1532 GAGCTCGAGAGCAGCTGAAAGCGCGTGCAGCAGCAGCTGTGCGCTGTCTCAGCGCCCA 1591
Qy      1612 CAGAACCAACCA-----AGAAAAAGAGAGAAAGCAGAGAGAGAGAGAGAGAGAGAG 1659
Db      1592 GTAAACAAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1651
Qy      1660 GAAAGAGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1719
Db      1652 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1711
Qy      1720 CCTCTAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1779
Db      1712 CCGCTGCCAGAGAGCTCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1771
Qy      1780 GCGCCCATGAG-----AGCAAGCCCGCTCCACGATGATGATGAG 1821
Db      1772 AGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1831
Qy      1822 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1881
Db      1832 GAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1891
Qy      1882 AACCAAGTCCCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1941

```

---

```

Db      1892 AACCGGCTCCCGGAGAGAGAGCTGGCGCGGAGTAGTCACATCATCAATCTCGGAGACC 1951
Qy      1942 TCCCTGAAGAAATTCACACCCGAGAGATGAATCGACTTTGAGACCTGAGCGCTCC 2001
Db      1952 TCGCTCAGAGAGACTCCACCCGAGAGAGATGAATGACTTTGAGACTGTGAACCCACC 2011
Qy      2002 AACTGCGTGAAGCTTGGAGCGCTATGTCACCTCTGTGTTGGGAGAGAGAAAGAACTTC 2061
Db      2012 ACTTGGCGGGAAC-TGGAGAGATATGTCAGACTGTGTTACAGAAAAGCAAGAGAAACC 2070
Qy      2062 AAGCTGAGAGAG 2073
Db      2071 GTTCTCAGCAAG 2082

```

---

```

RESULT 8
LOCUS   BC031536
DEFINITION Mus musculus, bromodomain-containing 3, clone MGC:28328
ACCESSION BC031536
VERSION   BC031536.1 GI:21594630
KEYWORDS MGC.
SOURCE   house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 3033)
AUTHORS Strausberg, R.
TITLE     Direct Submission
JOURNAL   Submitted (06-JUN-2002) National Institutes of Health, Mammalian
          Gene Collection (MGC), Cancer Genomics Office, National Cancer
          Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550,
          USA
REMARK   NIH-MGC Project URL: http://mgc.ncl.nih.gov
COMMENT   Contact: MGC help desk
          Email: cgabs-rtmail.nih.gov
          Tissue Procurement: Gilbert Smith, Ph.D.
          cDNA Library Preparation: Life Technologies, Inc.
          DNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNLT)
          Sequencing Center
          Center code: BCM-HGSC
          Web site: http://www.hgsc.bcm.tmc.edu/cdna/
          Contact: amg@bcm.tmc.edu
          Gunaratne, P.H., Garcia, A.M., Lu, X., Huiyk, S.W., Hale, S.M.,
          Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
          Richards, S., Gibbs, R.A.
          Clone distribution: MGC clone distribution information can be found
          through the I.M.A.G.E. Consortium/ILNLT at: http://image.llnl.gov
          Series: IRAC Plate: 36 Row: m Column: 24
          This clone was selected for full length sequencing because it
          passed the following selection criteria: matched mRNA gi: 12963646.
          Location/Qualifiers
            1. 3033
              /organism="Mus musculus"
              /db_xref="taxon:10090"
              /map="C2BCH.11"
              /clone="MGC:28328 IMAGE:4015879"
              /tissue_type="Mammary tumor metastasized to lung.
              MMTV-LTR/Int1 model. Expression driven by an MMTV-LTR
              enhancer."
              /clone_1lb="NCI CGAP_Lu30"
              /lab_host="DH10B"
              /note="Vector: pCMV-SPORT6"
              167..2347
              /codon_start=1
              /product="bromodomain-containing 3"
              /protein_id="AA031536.1"
              /db_xref="GI:21594631"
              /db_xref="LocusID:67382"
              /translation="MSTTAAPTGIPAVPGVPVNPPEVSNPSKPKRTKNTQLOVMONV

```



[illegible]

QY	151	CGCCCCCAGAGACTCCAAACCTTACAAAGCCCAAGAGCGACACCAACTGCATATC	210
Db	201	CCCCCACTGTGGGTCTCCAAACCCAGCAAGCCCGGGGGAAGAGCTAACCAACTGCAGTAC	260
QY	211	CTGCCTACAAGGAGGTCGCAAGACACTGTGGAAACACCAAGTTCATGAGGCTTCCAGGAC	270
Db	261	ATGCAGATGTAGTGGTGGAGACACTGTGAAACATCAGTTCCCTGGGCTTTTCAACAG	320
QY	271	CTGTGATGTGCGCTCAAGAGCTAAACCTCCCTGATTACTATTAAGATCTTAAACGGCTATG	330
Db	321	CCTGTGATGTGCAATCAAGAGCTAAACCTCCCTGATTATCATTAATTAATAAAAACCAATG	380
QY	331	GATATGGGAACAATTAAGAAAGCGCTTGGAAACAACTATTACTGAAATGCTCAGAAATGT	390
Db	381	GACATGGGGACTATTCAGAAAGAGACTGAAATAATTAATTATGTAGTGCCAGTGAAGTGT	440
QY	391	ATCCAGAGCTTCAACACTATGTATTACAAATTTGTACATCTACACAAAGCTTGGAGATAC	450
Db	441	ATGCAGGACTTCAACACCACTGTTTACAACTGTTATTAATTAATTAAGCCCAACATGAT	500
QY	451	ATATGCTTAATGGCAGAGAAGCTGGAAGAAACCTCTCTGCGAAAAATTAATAGCTATACC	510
Db	501	ATATGCTTAATGGCCCAAGGCTTATGAGAAATTTTCTGTGCAAAATGGCCCAAGATGCT	560
QY	511	ACAGAGAAACCGAGATCATGATAGTCCAGCAAAAGAGAGAGAGCTGGAGGAAGAA	570
Db	561	CAGGAGAAATGTGAAGTATTTGCCCCCTGCTCCAAAGGCGCAAAAGGCC - - GGAAGCAGGT	617
QY	571	ACAGGAGACGAAACCTGGCGTTTCCAGGGTACCAACACAACTGACATGACTCT	630
Db	618	GCAGAGAGCCAAATGTGACGGTTCCAACAAAGTGGACGCTGTGCTTCGCCAGCA	677
QY	631	CCGCGAGACCCAGACCCCTCAGAGCCGAATCTCTCTGTGAGGCGCAAGCGCTCAACCCCTTC	690
Db	678	ACCCCTTCGGAACATATACCCCA - - - - - CCCTGTCCAGACACCCGTC	722
QY	691	CTGCGCTCACCCCGGAGCTATCTGTCCAGACCCCTGTCAATGACAGTGGTCCCTCCAG	750
Db	723	ATTGCGCGTACCCCTGTATCCAAACATCATGTCAAAAGTCACTGACGTAC - - - - -	769
QY	751	CCACTGCGAGACGCCCGCGCAGTACCCCGCCAGCAGCAACCCCGCACCGCTCCAGCTCC	810
Db	770	-----TCCAGTCCCGCCACCCGCTCACCGCT	797
QY	811	CAGCCCGTACAGAGCCACCCACCATATGTGGGGCCACCCGACAGCTGTGAAGCAAG	870
Db	798	CTCTCTGCGACACCATCTGTCCTGTGTCTCTCCA - - - CACCGCTGTACTCAAGAA	854
QY	871	AAGGAGTGAAGAGAAAGCAGACACCAACCCCGCCACCAACATTGACCCCAATTAC - - -	927
Db	855	AAGGGCGTGAAGGGGAAAGCAGACACCAACCAACATCTGCTCATCTGCTCCAGC	914
QY	928	-----GAGCCACCTCTGCTGCCCCCGGAGCCCAAGACCAACCAAGTGGGCGAGCGG	981
Db	915	CGGAGTGAATCTCCCGCCACCACTTTCAGAGCCCAAGCAAGGTAAGTGGCCCGGAAG	974
QY	982	GAGAGC - - - - - AGCCGGCTGTGAAACCTCCCAAAAGAGAGTGGCGGACTCAGACAGC	1038
Db	975	GAGAGCGGGGCGCCCGCCATCAACCTCCCAAAAGAGACCTGGAGATGTGAGTCCCA	1034
QY	1039	CCAGCACAGAGAAAGACACAGAGTCTCGAGACAGCTCAAGTGTGACAGCGGCACTCTC	1098
Db	1035	CAGCACGGGGGTAAAGAGGAAAGTCTCGAGACCACTGGGCACTGTACAGCATCTCTC	1094
QY	1099	AAGAGATGTTTGGCAAGAAAGCAGCGCGCTACGCTGGCCCTTCTACAAGCTGTGCAC	1158
Db	1095	CGGGAGATGCTGTCCAAGAAAGCATGTGCTCTATGCTGGCCCTTTTACAAAGCACTGCAC	1154
QY	1159	GTGGAGCACTGGGCTACACAGCATCTGTGATATATCAAGAGCCCATGGACATGAGC	1218
Db	1155	GGGAGGCGCTGGAGCTGCAATGACTACATGACATATCAAGACACCCCATGGACCTCGAC	1214
QY	1219	ACAATCAAGTCTAAACTGGAAGGCCGAGTACCTGATGCTCAGAGATTTGTGCTCAC	1278



ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 2548)  
 AUTHORS Strausberg, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-JUN-2002) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 REMARK NIH-MGC Project URL: <http://mgc.ncl.nih.gov>  
 COMMENT Contact: MGC help desk  
 Email: [gcgaps-remail.nih.gov](mailto:gcgaps-remail.nih.gov)  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: National Institutes of Health Intramural  
 Sequencing Center (NISC),  
 Gaithersburg, Maryland;  
 Web site: <http://www.nisc.nih.gov/>  
 Contact: [nisc.mgc@nih.gov](mailto:nisc.mgc@nih.gov)  
 Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,  
 Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,  
 Dietrich, N.U., Granite, S., Guan, X., Gupta, J., Haghighi, P.,  
 Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,  
 Maduro, Q.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,  
 McDowell, J., Pearson, R., Stantirpop, S., Thomas, P.J., Touchman, J.W.,  
 Turegion, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,  
 Young, A., Zhang, L.-H. and Green, E.D.  
 Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAL Plate: 39 Row: 9 Column: 22  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 12408642.  
 FEATURES  
 source  
 1..2548  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="MGC:29612 IMAGE:4856840"  
 /tissue\_type="Colon, adenocarcinoma"  
 /clone\_lib="NIH-MGC\_15"  
 /lab\_host="DH10B-R"  
 /note="Vector: pOTB7"  
 308..1978  
 /codon\_start=1  
 /product="Similar to bromodomain containing 3"  
 /protein\_id="AAH32324.1"  
 /db\_xref="GI:21594670"  
 /translation="MSTATVPAAGIPAPPGVNPPEVSNPSKPGKTNLOLYMON  
 VVVKTLMKHOFAMPFYQVDAILNLDPYRIKINPMKGTIKRLNNYSGASECM  
 ODENITNICYIYNKPTDIDIVLAQALEKIFLOKVMQPEEELLPAPKGRKRA  
 AGASAGTOOVAVSSVSPATPROSVSPVIAATVPITANVSPVPPAA  
 PPPATPIVVPVPPVPPVKKKADPTPTATATASRESSEPPISDPQAVY  
 ARRESGRPIKPKPKLEDESEVYQHAKKKESEHLRYCSITIREMISKHNAKMPF  
 YKRVDALELHDYHDIKHPMDLSTYKRMDEKREYPDAGFAVDVRLMNSYKTPN  
 PDHEVVAAMARKLQDVEEMRPAPKMPDEVEPALPAPAPVAPKPKKKEKKEKKEK  
 GSSDSEERATRLAEQDLKAVHEQILASQAPVAPKPKKKEKKEKKEKKEK  
 EKHVKAEEKKKAKVAPPAQAOOKAPPAKANKSTTAAAGDHLTGV"  
 BASE COUNT 641 a 754 c 753 g 400 t  
 ORIGIN  
 Query Match 23.3%; Score 541.6; DB 9; Length 2548;  
 Best Local Similarity 63.2%; Pred. No. 6e-108;  
 Matches 1014; Conservative 0; Mismatches 494; Indels 96; Gaps 8;

QY 244 CACAGTTTGCATGGCTTTTCACAGAGCTGTGATGCGCTCAAGCTGACCTCCAT 303  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 464 CACAGTTTGCCTGGCTTCTTACAGAGCTGTGATGCGCTCAAGCTGACCTCCAT 523  
 QY 304 TACATTAAGATCAATTAAGACCTATGATATGGAAACATTAAGACCTGGAAAC 363  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 524 TATCATTAATAATTAATAAACCAATGATATGGGAGTATTAAGAGACTGAAAT 583  
 QY 364 AACATTAAGTGAATGCTCAGAGATGATATCAAGACTTCAACTATGTTTACAAAT 423  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 584 AATTATTAATTTGGATGCAAGGCAATGATATGCAAGCACTTCAACACTGTTTACAAAT 643  
 QY 424 TACATTTCAAGACCTGGAGATGATGATCTTAAATGGAGAGACTTGGAAAGCTC 483  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 644 TACATTTTACAAAGCCACAGATGATGATGATGATGATGATGATGATGATGAT 703  
 QY 484 TTTCTGCAAAATTAATGACTACCCACAGAGAAACCGAGATCATGATGATGATGAT 543  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 704 TTTCTGCAAAATTAATGACTACCCACAGAGAAACCGAGATCATGATGATGATGAT 763  
 QY 544 AAAGAGAGAGAGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 603  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 764 AAAGAGAGAGAGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 802  
 QY 604 CCAAGACAACTCAAGCATGCTCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 663  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 803 GCAGGTACAG 830  
 QY 664 CTTGTGACAG 723  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 831 CTTGTGTCTCCAG 866  
 QY 724 CTTGT 783  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 867 CCGGTCA-----TCGCTCCACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 937  
 QY 784 CCAGACCCACCCAG 843  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 938 GTCCAGT 997  
 QY 844 GCCACCCACAGCTGT 903  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 998 CTTGT 1057  
 QY 904 CCGA-----CCACATGACCCATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 954  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1058 CCGACAG 1117  
 QY 955 AAGACCAAGCTGT 1011  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1118 AAG 1177  
 QY 1012 AAG 1071  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1178 AAG 1237  
 QY 1072 CAGCTCAAGT 1131  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1238 CAGCTGT 1297  
 QY 1132 GCTGT 1191  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1298 GCTGT 1357  
 QY 1192 ATCATCAAG 1251  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1358 ATCATCAAG 1417  
 QY 1252 CGT 1311  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1418 CCAAG 1477







```

Db 1332 TCCGGAGCCTGTGATGCTCAAACTGGCTTACCGGTTATTCACAAAATTTATAAC 1391
OY 323 CGCCTATGATATGGAACAAATAAAGAGCGTTGGAAAAACAATTTACTGGATGTC 382
Db 1392 AGCCTATGGAACATGGGTACTATTAAAGGAGACTTGAACAAATTTATTTGGGCT 1451
OY 383 AGAATGTATCCAGGACTTCAACATATGTTTCAAAATTTTTCATCTACACAGCTG 442
Db 1452 CAAAGATATGCAAGATTTTAAATACATGTTCCAACTGTTCAATTTTAAACAGCCCA 1511
OY 443 GAGATGACATAGTCTTAATGGCAGAAAGCTGTGAAAAAGCTCTTGCAAAAAATAATG 502
Db 1512 CTGATATATTTGCTTAATGGCACAACGCTGAAAAAGATATTCTTACAGAAAGTTGCA 1571
OY 503 AGCTACCAACAGAAAGAACGAGATCATGATAGTCCAGGCAAGAAAGAGAGAGAGCTGGA 562
Db 1572 CATGCCACAGAAAGAAACAAGAGCTGTAAGTACATCCCTTAAGAACAGCACAAGAGG 1631
OY 563 GGAAGAAACAGGAGACAGCAAAACCTGGGTTTCCAGGGTACCACAAACACATCAAGCAT 622
Db 1632 GGGCCAAAGTGGCAGGCTCCAGGCGAGTGTACAGTGCCCATCAGTGCTGCTGCTGCT 1691
OY 623 CGACTCTCCGAGACCCGAGCCCTCAGCCGAAATCTCTCTGTCAGAGGCGCAGCTC 682
Db 1692 -----CTTCTGTGTACACACAGCCCTGTATATCTCTCAGCTGAGATACCTACACTG 1745
OY 683 ACCCTTCCCTGCGCTACCCGAGCCTCATGCTCCAGACCCCTGTATGATGAGTGGTGC 742
Db 1746 TCCTCACATVCCCAACCATCATGCTATTTCC-----T 1778
OY 743 CTCCTCCAGCATGAGAGAGCCCGCCGAGTGGCCCCCAAGCACAACCCCAACCCGCTC 802
Db 1779 CTCACATTTCTCAAGCTCTTGACATCTGTGAGACCCCGCTCTCTGCTGTTACAGCTGC 1838
OY 803 CAGCTCCAGCCGCTGACAGAGCCACCAACCATCATGCGGCGGCGCCACAGCCTGTGA 862
Db 1839 CTCACAGCCAGCCCTTGGCAGAAAAAAGGCT----- 1872
OY 863 AGACAAAGAGGAGTGAAGAGAAAGACAGACACCAACCCCAACCCCAACCATTTGACCCCA 922
Db 1873 ----AAAGGGAAGAGAGTACTACCAACCCCTACACCTACAGCCTATTTGGCTCTGTT 1928
OY 923 TTCAGAGCACCCTGCTGCGCCCGGAGCCCAAGACACCAAGCTGGGCG-----ACGGC 979
Db 1929 CTCACACTAGCCCTCTGAGAGTCTTGAGCCTTAAGGACAGCGCTTCCCTATCCGTA 1988
OY 980 GGGAGAGCAGCGGCTGTGAAACCTCAAAAGAGAGCGTCCGAGCTCTCAGCAGCAGC 1039
Db 1989 GAGAGAGTGGTGGCCCATCAAGCCCCACAGCAAGACTTGGCTGACTCTCAGCAACAC 2048
OY 1040 CAGCACCAGAGAGAGCAGCAGAGCTCGAGAGCCTCAAGTCTCAGCGGCATCTCA 1099
Db 2049 ACCAGAGCTCTAAGAAAGAAAGCTTTCAGAACAGTTAAACATTCATGGAATTTTGA 2108
OY 1100 AGAAGATGTTTGGCAGAGAGCAGCGCGCTAGCGCTGGGCTTCTCAAGGCTGTGGAGC 1159
Db 2109 AGGATTAATCTCTTAAGAGCAGTGTGCTTGGCTTCTTAAACCAAGTGGAGT 2168
OY 1160 TGGAGCAGTGGGCTTACAGCAGTCTGATCATCATCAAGCAGCCCATGAGATGAGA 1219
Db 2169 CTTTGCATTTGGCTTGCATGATCATCATCATTTAAGCAGCCCATGAGACTTACGA 2228
OY 1220 CAATCAAGTCTAACTGAGAGCGCGTGAAGTACCGTATGCTCAGAGATTTGCTGAGC 1279
Db 2229 CTGTCAAGGGAAGATGAGAACCGTGAATACGGGATGACAGAGATTTGGCTGAGT 2288
OY 1280 TCGGATGATGTTCTCAACTGCTATAAGTACAAACCTCTGACATGAGGTGGTGGCA 1339
Db 2289 TACGGTGTATGTTCTCAACTGCTATAAGTACAAACCTCTGACATGAGTGTGGCA 2348
OY 1340 TGGCCGCAAGCTCAGATGTTGTAATGCGCTTTCGAAGATGCGGAGAGAGCAGCTG 1399

```

```

Db 2349 TGGCAGAAAGCTACAGAGATGATTTAGTTCGGTTATGCCAAGATCCAGATGAACAC 2408
OY 1400 AGAGCCAGTGTGTGGCCGCTGCTCCCGGCAAGTGGCCCTCC-----ACCAAGTTGTG 1456
Db 2409 TAGAACAGGCGCTTTACAGAGTCTTACGTCCATGCCCCCTGGCTTGGCAATTCGTT 2468
OY 1457 CCCCCCTTCATCCAGCAGCAGCAGCAGATGCTCTC----- 1496
Db 2469 CAGAGTCTTCAGTGAAGAAAGTACAGTGAAGAGCTCTCTGAGAAAGAGAGAGAAG 2528
OY 1497 -----GCAAGTGAAGTTCGATGATACCTGAGCTGAGGAGAGAGAGAGCC 1540
Db 2529 ATGAGAGAGCAGAGAGAGAAAGAGAGTGAAGAGTCAAGCTTCAAGAGAGAGAGAGG 2588
OY 1541 AGCGCTGTGCTGAGCTTCCAGAGCAGCAGTCAAAAGCCGTGACAGAGAGCTTGCAGCCCT 1600
Db 2589 ATGCTTGAAGAGATACAGAGAACACTTGGGCACTACATGACAACTGGCTGCTGT 2648
OY 1601 CTGAGCCCAAGCAAGAAACCAAGAG-AAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1659
Db 2649 CCCAGGCTCCATATATCAAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2708
OY 1660 GAAAGCAGCAAAAGAGAGAG-----GAAGTGAAGAGATTAATTAAGAGAGAGAGAG 1705
Db 2709 GGAAGGCAAGAGAGATCGAGGCGAGCTGGGCGCAGCAGAGATGACAAAGGGGCTA 2768
OY 1706 AAGCCAAAGAACCTCTCTTAAAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1749
Db 2769 CACCCGCGCCACTCAACCTTAAGAGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2828
OY 1750 -----AGCAACAGCAATGTGAGCAGAG 1771
Db 2829 CTGCTTAAAGCCCTTCTGCTTGTGACCTTCTGAGAGAGTGGCAGCAAGCTCCCAAA 2888
OY 1772 AGAGAGCAGGCGCCATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1831
Db 2889 AGGCACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2948
OY 1832 AGTCAAGCCTATGCTCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1891
Db 2949 AGAGCAGGCGCCATGAGTATGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3008
OY 1892 CCGGCGAGAGAGTGGGCGCGCTGTGATCATCATCATGACAGAGAGAGAGAGAGAGAG 1951
Db 3009 CTGGGAGAGAGAGTGGGCGAGTGTGATATATCAACAGAGAGAGAGAGAGAGAGAG 3068
OY 1952 ATTCCAAAGCCGAGAGATTTGAATGAGCTTGAAGAGAGAGAGAGAGAGAGAGAGAG 2011
Db 3069 ATTCAAGCCAGAGAGATTTGATGATTTGAAACAGTCAAGAGAGAGAGAGAGAGAG 3128
OY 2012 AGCTTGAAGCCTATGATCACTCTCTGTTTGCAGAGAGAG 2050
Db 3129 AGCTT-GAGCGGTATGCTTCTCTGCTTACGTAAGAGAA 3166

RESULT 14
HUMFSHG 4053 bp mRNA linear PRI 22-JUN-1995
LOCUS Human homolog of Drosophila female sterile homeotic mRNA, complete
DEFINITION cds.
ACCESSION M80613
VERSION M80613.1 GI:182768
KEYWORDS MHC class II antigen; female sterile homeotic gene.
SOURCE Homo sapiens CDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS Beck,S., Hanson,J.M., Kelly,A., Paplin,D.J.C. and Trowsdale,J.
TITLE A homologue of the Drosophila female sterile homeotic (FSH) gene in
the class II region of the human MHC
JOURNAL DNA Seq. (1991) In press
REFERENCE 2 (sites)

```



AUTHORS Haynes, S.R., Dollard, C., Winston, F., Beck, S., Trowsdale, J. and David, I.B.  
TITLE The bromodomain: a conserved sequence found in human, *Drosophila* and yeast proteins  
JOURNAL Nucleic Acids Res. 20 (10), 2603 (1992)  
MEDLINE 92285152  
PUBMED 1350857

FEATURES  
SOURCE Location/Qualifiers  
1..4053  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_type="T CELL"  
1178..1342  
/note="putative"  
/codon\_start=1  
/protein\_id="AA68890.1"  
/db\_xref="GI:182769"  
/translation="MASVPAQLQLTPANPPPEVSNPKPRGVNTQLOYLKHYKALM  
KHOPAMPFROPVAVKLGPDYHKTIKOPMDGTRKRLNNYMAEBCMODNTWF  
TNCYIYNKPTDDIVLMAQLEKIFLOKVASMPQEOELVVTIPKNSHKAKLALOG  
SVSAHOVPVAVSSVSHALYTPPEIPTVLIPIHPVISPLKSHSGAPLLAAT  
AAPPAOLAKKGVKRAADYTPPTAILAGSPASPGSLPEKARLPPMRSEGRP  
IKPPKRLDPSOOHOSKKKLSSEOLKNGILIKELSKHAAVAMPKPVDAAL  
GLHDYDIITKPMPLSPVKRKMNRDVRDQBPADYRLMFNCKYKNPPDHVYAL  
RKLDYDFEPRYAKMDEPLEPLVSTAMPGLAKSSSESSSESSSESESEEE  
DEDEDESESSSDSEERARLAELOEQLRAVHQLAALSOGPISKPKRREKREK  
KKRKAHEHRGADDEDKRAPRAPPOPKSKKASGSGSALGSGFSGSGST  
KLPRKATKTPALPTGYDSESEESRPMYDEKROLIDINKLPGLGRVILIOA  
REPLRDSNPEELIDPELTLPKPLRELYELSECLKRKRKPYTIKKPKVKEELA  
LEKKRELEKRLDQVSGOLNTRKPRKANKESSSSAQVAVSHLSASSSSSSSSSS  
SSSSSDTSDSDSG"  
4015..4020  
polyA\_signal  
/note="putative"  
BASE COUNT 997 a 1132 c 1067 g 857 t  
ORIGIN

Query Match 19.8%; Score 460.2; DB 9; Length 4053;  
Best Local Similarity 57.0%; Pred. No. 4.5e-90;  
Matches 1151; Conservative 0; Mismatches 693; Indels 175; Gaps 10;

QY 143 CCAACCCCCCGCCGAGAGCTCCAAAGCCCTAACAAAGCCCAAGAGGAGAGCAACCAAC 202  
DB 1212 CCAACCCCAACACCCCGAGAGTGTCCAAATCCCAAAAACCGAGAGATTACCAACCGC 1271  
QY 203 TGCATTAAGTCTGCTGAGAGTGTCTCAAGACATATGGAACACACCTTTCATGAGCTT 262  
DB 1272 TGCATTAAGTCTGAGAGTGTCTGAGAGTGTCTGAGAGTGTCTGAGAGTGTCTGAG 1331  
QY 263 TCCAGCAGCTGTGAGTGTCTGAGAGTGTCTGAGAGTGTCTGAGAGTGTCTGAGAGT 322  
DB 1332 TCCAGCAGCTGTGAGTGTCTGAGAGTGTCTGAGAGTGTCTGAGAGTGTCTGAGAGT 1391  
QY 323 CGCCTATGATATGGAAGCAATTAAGAGCGCTTGGAAGAAACACATATTACGAGATGCTC 382  
DB 1392 AGCCTATGAGATGGAAGCAATTAAGAGCGCTTGGAAGAAACATATTATTTATGAGCTG 1451  
QY 383 AGAATGATATGAGAGCTTCAACATATGTTTACAAATTTTACATGTAACAAGCTG 442  
DB 1452 CAGAGTGTATGAGAGTATTTAATACATGTTACCAACTGTTACATTTACAAAGCCCA 1511  
QY 443 GAGATGATATGTTTAAATGAGAGCTGTGAAAGAGCTTCTTGCAGAAATTAATG 502  
DB 1512 CTGATGATATGTTTAAATGAGAGCTGTGAAAGAGCTTCTTGCAGAAATTAATG 1571  
QY 503 AGTACCAACAGAGAAAGCAAGATCATGATGATGAGAGCAAGAAAGAGAGAGTGGGA 562  
DB 1572 CAATGCCAACAAGAAAGCAAGATGATGATGATGATGATGATGATGATGATGATGATG 1631  
QY 563 GGAAGAAAGAGAGAGCAAGAAAGCTGCGCTTTCACAGGTACCAAGCAACAATCAAGAT 622  
DB 1632 GGGGCAAGGTGAGAGAGCTGCAAGAGAGTGTACCAAGTGGCCATCAGAGTGTCTGCT 1691  
QY 623 CGACTCTCCGAG 682

DB 1692 -----CTTCTGTGTGACACACAGAGCCCTGATATACCTCTCCACAGAGATACCTACACTG 1745  
QY 683 ACCCTTCCCTGCGGTGACACCCCGAGACCTCATGCTGCACACCCCTGTATGACAGTGTGC 742  
DB 1746 TCTCTACATTTCCACCCCATGATGATTTCC-----T 1778  
QY 743 CTCCCAACCACTGACAGAGCCCGGAGAGTCCCGCCCGCCAGCACACCCCGCCGCTC 802  
DB 1779 CTCCACTCTCAAGTCTTGTGACACTGTGTGAGAGCCCGCCCTCTTGTCTTACTGAGCTC 1838  
QY 803 CAGCTCCCAAGCCGATACAGAGCAACCCATCATATCTGCGGACCCACAGCCCTGTGA 862  
DB 1839 CTCCAGCCAGCCCTTCCCAAGAAAGAGCGT----- 1872  
QY 863 AGACAAAG 922  
DB 1873 -----AAAGCGAAGAGAGATATCTACACACCTTACACCTTACAGCCATCTTGCTCTGCTT 1928  
QY 923 TTCAAGAGCAACCTTCTGCTGCGCCCGGAGAGCCCAAGAGACCAAGCTGGGCGC ---AGCGC 979  
DB 1929 CTCCAGTACGCTCTCTGAGAGTGTGTGAGCTTAAGGAGAGAGAGAGAGAGAGAGAGAG 1988  
QY 980 GGGAG 1039  
DB 1989 GAG 2048  
QY 1040 CAGACAG 1099  
DB 2049 ACCAGAGCTTACAG 2108  
QY 1100 AG 1159  
DB 2109 AG 2168  
QY 1160 TGAAG 1219  
DB 2169 CTCTGAG 2228  
QY 1220 CAATCAAGTCTTAACTGAG 1279  
DB 2229 CTGTCAAG 2288  
QY 1280 TCCGATGATGTTCTCCAACTGCTATATAGTACAAACCTCTGAGACATGAGTGTGGCA 1339  
DB 2289 TACGGCTATTTCTCCAACTGCTATATAGTACAAACCTCTGAGACATGAGTGTGGCA 2348  
QY 1340 TGGCCCGAGAGCTCAGAGATGTTCGAATGCGCTTGGCCAGAGAGAGAGAGAGAGAG 1399  
DB 2349 TGGCAG 2408  
QY 1400 AGGAG 1456  
DB 2409 TAGAAG 2468  
QY 1457 CCGCCGCTTCTCCAG 1496  
DB 2469 CAGAGTCTCCAG 2528  
QY 1497 -----GACAGTGAAG 1540  
DB 2529 ATGAG 2588  
QY 1541 AGGAG 1600  
DB 2589 ATGCTTATGAG 2648  
QY 1601 CTGAGCCCAAG 1659  
DB 2649 CCGAG 2708  
QY 1660 GAAAG 1705

Db	2709	GGAAAGCGAGAGAAAGCATCGAGGCCGACGCTGGGGCCGACGAGATGACAAAGGGGCGTTAGGG	2768
QY	1706	AAGCCAGAGAAACCTCTCTCTTAAAAAAGACGACAAAATATATGCG-----	1749
Db	2769	CACCCGCCACCTCAACCTAAGAAAGTCCAAAGAAAGCAAGTGGCAGTGGCGACGTG	2828
QY	1750	-----AGCAACGACAAATGTGACGACAA	1771
Db	2829	CTGCTTTAGGCCCTTCTGGCTTTGGACCTCTTCGGAGGAAGTGGCACCAAGCTTCCCAAAA	2888
QY	1772	AGGAGCCAGGCGCCATGAAGAGCAAGCCCCCTCCACGTATAGTGGAGGAAGAGACA	1831
Db	2889	AGGGCCACAAAGACACCCCCACCTGCGCTTCACAGGTATATGCAGAGGAGAGGAAG	2948
QY	1832	AGTGAACGCTATGTGCTATGAGGAGGAAGGGCGACCTCAGCTTGGACATCAACAAGCTCC	1891
Db	2949	AGAGGAGGCGCCATGAGTTATCGATGTAGAAAGCGGCACCTGTGCGTGGACATCAACAATATAC	3008
QY	1892	CCGGCGAGAGAGCTGGGGCGCGCTGTGTCACATCATCGATCGAGGAGAGGCCCTCCCTGAAAGA	1951
Db	3009	CTGGGGAGAGAGCTGGGGCCGAGTGTGTCATATATATCCAAAGCCAGGGAGGCCCTTTTACGTG	3068
QY	1952	ATTCCACCCCGACAGATTTGAATTCGACTTGTGAAGACCCGTGAAGCGCTGCACACTGCGCTG	2011
Db	3069	ATTCAACCCCGAAGAAAGATTGAGATTGATTGTGAACACTCAAGCCATCAACACTTAGAG	3128
QY	2012	AGCTTGAGCGCTATGTCACTCCCTGTTTGGCGAAGAAA	2050
Db	3129	AGCTT-GAGCGCTATGTCCCTTCGCGCCATCTTAAGAAA	3166

RESULT 15			
AX409010			
LOCUS	AX409010	4664 bp	DNA
DEFINITION	Sequence 1657 from Patent WO0229103.		linear
ACCESSION	AX409010		
VERSION	AX409010.1	GI:21441715	

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1  
AUTHORS Alvaras,C., Horne,D., Peres-da-Silva,S. and Vockley,J.G.  
TITLE Gene expression profiles in liver cancer  
JOURNAL Patent: WO 0229103-A 1657 11-APR-2002;

FEATURES	Location/Qualifiers
source	1. .4664

	/note="EMBL/GenBank Accession No. D42040			
BASE COUNT	1084	a	1336	c 1293 g 951 t
ORIGIN				

Query Match	19.8%	Score 460.2	DB 6	Length 4664
Best Local Similarity	57.0%	Pred. NO. 4.6e-90		
Matches 1151; Conservative	0	Mismatches 693	Indels 175	Gaps 10

0Y	143	CCACCCCCCGCCCCCGAGACCTCCACACCTTCAACAAGCCCAAGAGCAGACCAACCAAC	202
Db	1877	CCAAACCCACCAACCCCGAGAGTGTCCAAATCCCAAAAAGCCAGGACGATTTACCAACGAC	1936
0Y	203	TGCATATACCTGCTCAGAGTGGTCTCAAGCACTATGAGAAACACCGATTGGATGGCTT	262
Db	1937	TGCATATACCTACACAAAGGTAGTATGAAAGGCTCTGTGGAACATCACTGTCGATGGCCAT	1998
0Y	263	TCCAGCAGCCTGTGGATGCGCGTCACAGTGAACCTCCCTGATTCATTAAATCATTTAAA	322
Db	1997	TCCGCGACGCTGTGGATGCGTCAAACTGGGTTACCGGATTTATCACAAAATTAATAAAC	2056
0Y	323	CGCCTATGAGATGGGAAACAATTAAAGAGCGCTTGGAAAACAATTACTGGAAATGCTC	382

D	2057	AGCGTAGGACATGGGTACATTTAAGAGGAGACTTGAAAACCATTTATTTTTGGGCTGCTT	2116
Q	383	AGGAATGTATTCAGGACTTCAACATATGTTTACAAATTTGTATCATCAACAAGCTTG	442
D	2117	CAGAGTGTATCAAGAAATTTTAATACCATGTGTTACCAACTGTATTATTTTCAACAAGGCCA	2178
Q	443	GAGATGACATAGTCTTAATGGCAGAAAGCTGTGGAAAAGCTCTTCTTGCAAAAATTAATG	502
D	2177	CTGATGATATTGTCTTAATGGCAACAAAGCTGGAAAAAGATATTTCTCAAGAAAGTTTGAT	2236
Q	503	AGCTTACCACACGAAGAAACCGAGATCATGTATGTCCAGACAAAGAAAGAAAGAGACCTGGGA	562
D	2237	CAATGTCCCAAGAAAGAAAGAAAGAGCTGGTAACTAGACATCTCCTTAAGAAGACGCCAAGAAAG	2296
Q	563	GGAAGAAACACAGGACAGCAAAACCTGGCGTTTTCACGGTACCAAAACACACTCAAGCAT	622
D	2297	GGGGCCAATTTTGGCAGCCCTCCAGGGCAGTGTATACCATGGCCCATCAGGGCTCGCGCTG	2356
Q	623	CGACTCCGCGGCACAGCCAGACCCCTTAGCGGAATCTCTCTCTGTGACGGCAGCGCTC	682
D	2357	-----CTTGTGTGTCAACACAGACCCCTGTATCTCTCTCACCCTGAGATACCTTACCCTG	2410
Q	683	AACCCCTTCCCTGCGGTACCCCGGAGCTCTATGCTCCAGAACCCCTGTCATGACAGATGTTGC	742
D	2411	TCTCTACATTTCCCAACCCCATCATGATTTTC-----T	2443
Q	743	CTCCCCAGCCCATGACAGACGCCGCCCGCAGTGGCCCCCAGCACAAACCCACCCGCTC	802
D	2444	CTCCACTTCTCAATGCTCTTGCACACTCTGTGAGCCGCCGCTCTGTGCTGTATCTGACAGTC	2503
Q	803	CAGCTCCCAACCCCGTACAGACGCCACCCATCCATCTCATGGCGGCACCCCAAGCCTGTGA	862
D	2504	CTCCAGCCCAACCCCTTGGCCCAAGAAAAAGGCT-----	2537
Q	863	AGACAAAGAAAGGAGTGAAGAGAAAGAACACACACACACCCACACATTTGACCCCA	922
D	2538	-----AAAGCGGAAAGCAGATTACTACACACCCCTTACACTTACAGCATCTTGGGCTCCTGGTT	2596
Q	923	TTTCAGGACCAACCTCGCTGCCCCCGGAGGCCAAGACCAACAGCTGGGCG--AGCGGC	979
D	2594	CTCCAGCTAGCCCTCTCTGGGAGTCTTGAGGCGTTAAGGCACACAGGCTTCCCTTATGCGTA	2653
Q	980	GGGAGAGACGCGGCGCTGTAAACCTCCAAAGAAAGAGCTGCCCACTCTCAGCAGCAC	1039
D	2654	GAGAGAGTGTGCGCCCATCAAGCCCAAGCAAAACCTGGCTGCTCTCAGCAACAC	2713
Q	1040	CAGACACAGAAAGACGACGACGATCTCGAGCAGCTCAAGTGTGCTCAGCGACCTATCACA	1099
D	2714	ACCAAGAGCTCTAATAAAGAAAGACTTTCAAGACAGTAAACATTTGCAATGGCATTTTGA	2773
Q	1100	AGGAGATGTTGGCCAAAGACGACGCGCCTACGCTGGCGCTTTCACAAAGCCTGTGGAG	1159
D	2774	AGGAGTATCTCTTAAGAAACATGCTGCCCTATGCTTGGCTTCTTATAAACCAAGGATG	2833
Q	1160	TGAGAGCACTTGGGCTTACAGACTACTCTGTACATCATCAAGCACCCCATGACATGACGA	1219
D	2834	CTTCTGCACTTGGCTCTCATGACTTACCATGACATCATTTAAGACCCCAAGCACTCACA	2893
Q	1220	CAATCAAGTCTAACTGTGAGGCGCCGAGTACCGGTATGCTCAGAGGTTTGGTGTGAG	1279
D	2894	CTGTCAAGCGGAAGATGGAACACCGTGTATTTACCGGGATGACACAGAGTTTGTCTGCTATG	2953
Q	1280	TCCGATGTATGTTTCCCACTGCTATTAAGTACAAACCTCTGACCATGAGTGGTGGGCA	1339
D	2954	TACGGCTTATGTTTCCCACTGCTATTAAGTACAAATCCCCCAATCACAATGTTTGTGGCAA	3013
Q	1340	TGGCCCGCAGCTCTCAGGATGTGTGAAATGCGCTTTGCCAAGATGCCGAGCGACCTG	1399
D	3014	TGGGACAGAAAGCTCAAGAGATGTTATTTGAGTTCGGTTATGCCAAGATGCCAGATGAACAC	3073
Q	1400	AGGAGCAGTGTGTGGCGTGTCTCCGCCCGGCAAGTGGCCGCTGCC--ACCAAGTGTGGG	1456
D	3074	TAGAACCAGGGGCTTTTACACAGTCTCTATGTCACATGCGCCCTGGCTGGCGCAATGCTTT	3133



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: March 1, 2003, 11:50:44 ; Search time 480 Seconds  
(without alignments)  
10926.884 Million cell updates/sec

Title: US-09-700-590A-101

Perfect score: 2329  
Sequence: 1 gcagggatcactagcatgctc.....ngccttaaaaggtlnccca 2329

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapept 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

N\_Geneseq\_101002:.\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2316	99.4	2329	21	AA56719 Human transmembran
2	559.4	24.0	667	22	AA503029 Human diagnostic a
3	460.2	19.8	4664	24	ABR83660 Human CDNA differe
4	460.2	19.8	4664	24	ABN95159 Gene #1657 used to
5	371	15.9	1884	22	AAH18591 Human cDNA sequenc
6	294.2	12.6	3104	19	AAV68343 Transcriptional re
7	294.2	12.6	3106	20	AAK40172 WO9904265 Seq ID N
8	227.8	9.8	1862	23	ABY25740 Human prostate exp
9	224	9.6	424	22	AA526606 Human CDNA encodin

10	224	9.6	746	22	AA526193 Human cDNA encodin
11	208.2	8.9	1741	24	ABK84162 Human CDNA differe
12	207	8.9	352	22	AAE66021 Novel human polynu
13	203.8	8.8	6776	23	ABL04337 Drosophila melanog
14	168.8	7.2	209	20	ABV7340 EST clone BW2. Ho
15	159.4	6.8	560	24	ABO60244 Human colon cancer
16	159.2	6.8	450	21	AAZ80432 Human colon cancer
17	126.6	5.4	559	21	AAA43730 Mouse secreted exp
18	120.8	5.2	642	20	AAK29128 Polyucleotide RTP
19	92.8	4.0	597	23	ABV57572 Human prostate exp
20	85.2	3.7	807	23	ABL19177 Drosophila melanog
21	85.2	3.7	2807	23	ABL19176 Drosophila melanog
22	85.2	3.7	5017	23	ABL15790 Drosophila melanog
23	84.4	3.6	372	21	AAAC5986 Eucalyptus grandis
24	79.2	3.4	140	19	AAV68386 Human BAZ gene pri
25	76	3.3	10629	22	AAH28279 - Nucleotide sequenc
26	75	3.2	2175	23	ABL12585 Drosophila melanog
27	75	3.2	4229	23	ABL12584 Drosophila melanog
28	75	3.2	10598	23	ABL27241 Drosophila melanog
29	75	3.2	15667	23	ABL27240 Drosophila melanog
30	71	3.0	1542	23	ABL23049 Drosophila melanog
31	71	3.0	3542	23	ABL23048 Drosophila melanog
32	69.8	3.0	1445	21	AAK47126 Arabidopsis thalia
33	69.6	3.0	2247	21	AAK6949 Arabidopsis thalia
34	68.6	2.9	2442	21	AAK50976 Arabidopsis thalia
35	68.4	2.9	306	21	AAAC56643 Eucalyptus grandis
36	68.4	2.9	348	21	AAAC5981 Eucalyptus grandis
37	68.2	2.9	693	23	AA574240 DNA encoding novel
38	68.2	2.9	693	23	AA590715 DNA encoding novel
39	68	2.9	1686	16	AAO87587 DNA encoding leuco
40	67.8	2.9	9443	23	ABL16121 Drosophila melanog
41	67	2.9	819	23	AA568928 DNA encoding novel
42	66.8	2.9	441	23	AA575453 DNA encoding novel
43	66.6	2.9	2443	24	ABO54948 Human ovarian anti
44	66.6	2.9	2444	21	AACT7905 Human cancer assoc
45	66	2.8	642	23	AA571152 DNA encoding novel

## ALIGNMENTS

RESULT 1	
AA256719	AA56719 standard; cDNA; 2329 BP.
XX	AA56719;
AC	23-MAR-2000 (first entry)
XX	
DT	Human transmembrane protein HTMPN-22 encoding CDNA.
DE	
XX	
KW	Human: transmembrane protein; HTMPN; diagnosis; immunospecific;
KW	antiproliferative; neutroprotective; immune disorder;
KW	reproductive disorder; smooth muscle disorder; neurological disorder;
KW	gastrointestinal disorder; developmental disorder;
KW	cell proliferative disorder; ss.
XX	
OS	Homo sapiens.
XX	
FN	WO9961471-A2.
XX	
PD	02-DEC-1999.
XX	
PF	28-MAY-1999; 99WO-US11904.
XX	
PR	29-MAY-1998; 98US-0087260.
PR	02-JUL-1998; 98US-0091674.
PR	02-OCT-1998; 98US-0102954.
PR	24-NOV-1998; 98US-0109869.
XX	
PA	(INCY-) INCYTE PHARM INC.
XX	
PI	Tang YT, Lal P, Hillman JL, Yue H, Guegler KJ, Corley NC;

PI Bandman O, Patterson C, Gorgone GA, Kaser MR, Baughn MR:  
 PI Au-Young J;  
 XX WPI; 2000-072605/06.  
 DR P-PSDB; AAY57898.  
 XX  
 PT Proteins, polynucleotides, vectors, host cells and antibodies used to  
 PT diagnose, treat or prevent immune, reproductive, smooth muscle,  
 PT neurological, gastrointestinal, developmental and cell proliferative  
 PT disorders -  
 PS Claim 9; Page 191; 229pp; English.  
 XX  
 CC AA256698 to AA256776 encode AAY57877 to AAY57955 which represent human  
 CC transmembrane proteins designated HTPN-1 to HTPN-79, respectively.  
 CC The transmembrane proteins have immunospecific, antiproliferative and  
 CC neuroprotective activities. The human transmembrane proteins,  
 CC polynucleotides encoding them and other compositions and methods from  
 CC the present invention, can be used for the diagnosis, treatment or  
 CC prevention of immune, reproductive, smooth muscle, neurological,  
 CC gastrointestinal, developmental and cell proliferative disorders. The  
 CC HTPN's can be used to treat or prevent disorders associated with a  
 CC decreased expression or activity of HTPN.  
 XX

Sequence 2329 BP; 674 A; 691 C; 569 G; 382 T; 13 other:

Query Match 99.4%; Score 2316; DB 21; Length 2329;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGGATCTAGCATGTGTGGGAGAGGCGCCCTGGGAGGAGATTGAGAAATCTGCCA 60  
 DB 1 GCAGGATCTAGCATGTGTGGGAGAGGCGCCCTGGGAGGAGATTGAGAAATCTGCCA 60  
 QY 61 GTAATGGGGATGATAGATACTTCCCAATGTCTACACACAGAGGCCGCCCAACC 120  
 DB 61 GTAATGGGGATGATAGATACTTCCCAATGTCTACACACAGAGGCCGCCCAACC 120  
 QY 121 CAGCCAGCAACGCGACGACCAACCCCGCCCGCCAGAGACCTTCAACCTTAACAG 180  
 DB 121 CAGCCAGCAACGCGACGACCAACCCCGCCCGCCAGAGACCTTCAACCTTAACAG 180  
 QY 181 CCCAAGGCGAGCAACCAAGTCAATNCTGTCAGATGGTGTCAAGACATATG 240  
 DB 181 CCCAAGGCGAGCAACCAAGTCAATNCTGTCAGATGGTGTCAAGACATATG 240  
 QY 241 AAACACAGTTTGGATGCGCTTTCAGACGCTGTGGATGCCCTCAAGCTGAACCTCC 300  
 DB 241 AAACACAGTTTGGATGCGCTTTCAGACGCTGTGGATGCCCTCAAGCTGAACCTCC 300  
 QY 301 GATTACTATAGATCAATTAAGCGCTTATGATATGGAACAATAAGACGCTTGAA 360  
 DB 301 GATTACTATAGATCAATTAAGCGCTTATGATATGGAACAATAAGACGCTTGAA 360  
 QY 361 AACCACTATTACGGAATGTCAGAAATGATACAGACTTCAACCTATGTTTAAAT 420  
 DB 361 AACCACTATTACGGAATGTCAGAAATGATACAGACTTCAACCTATGTTTAAAT 420  
 QY 421 TGTTCATCTACCAAGAGCTGAGATGACATAGTCTTAATGACAAAGCTGTGAAAG 480  
 DB 421 TGTTCATCTACCAAGAGCTGAGATGACATAGTCTTAATGACAAAGCTGTGAAAG 480  
 QY 481 CTCCTCTGCAAAAAAATAATAGAGTACCCACAGAAAGAAACGAGATCATATGTCAG 540  
 DB 481 CTCCTCTGCAAAAAAATAATAGAGTACCCACAGAAAGAAACGAGATCATATGTCAG 540  
 QY 541 GCAAAAGGAGAGAGCTGGGAGGAAAGAAACAGGACAAACCTGCGCTTCCAG 600  
 DB 541 GCAAAAGGAGAGAGCTGGGAGGAAAGAAACAGGACAAACCTGCGCTTCCAG 600  
 QY 601 GTACCAAAACAACTCAGATGACCTGCGCAGAGACCCGCTCAGCGGAATCC 660  
 DB 601 GTACCAAAACAACTCAGATGACCTGCGCAGAGACCCGCTCAGCGGAATCC 660

QY 661 CTTCTGTGAGGCGCAAGCCTCAACCCCTTCTGCGGTCAACCCGAGCTATGTCAG 720  
 DB 661 CTTCTGTGAGGCGCAAGCCTCAACCCCTTCTGCGGTCAACCCGAGCTATGTCAG 720  
 QY 721 ACCCTGTATGACAGTGTGCTCCCAACGACCTGACAGCGCCCGGACATGCCCC 780  
 DB 721 ACCCTGTATGACAGTGTGCTCCCAACGACCTGACAGCGCCCGGACATGCCCC 780  
 QY 781 CAGCACAACCCCGACCGCTCCAGTCCCGACCTGACAGAGCCACCCATCATC 840  
 DB 781 CAGCACAACCCCGACCGCTCCAGTCCCGACCTGACAGAGCCACCCATCATC 840  
 QY 841 GCGGCAACCCCGACCGCTCCAGTCCCGACCTGACAGAGCCACCCATCATC 900  
 DB 841 GCGGCAACCCCGACCGCTCCAGTCCCGACCTGACAGAGCCACCCATCATC 900  
 QY 901 ACCCCACCAACCAATGACCCATTCACAGACCTCCTGCTGCGCCGAGCCCAAGCC 960  
 DB 901 ACCCCACCAACCAATGACCCATTCACAGACCTCCTGCTGCGCCGAGCCCAAGCC 960  
 QY 961 ACCAGCTGGGCGAGGCGGAGAGACAGCGGCTGTGAACCTCCAAAGAGAGCTG 1020  
 DB 961 ACCAGCTGGGCGAGGCGGAGAGACAGCGGCTGTGAACCTCCAAAGAGAGCTG 1020  
 QY 1021 CCGAGCTTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1080  
 DB 1021 CCGAGCTTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1080  
 QY 1081 TGTGCAAGGAGATCTCTCAAGAGATGTTTCCAAAGACAGCGCCCTACGCTG 1140  
 DB 1081 TGTGCAAGGAGATCTCTCAAGAGATGTTTCCAAAGACAGCGCCCTACGCTG 1140  
 QY 1141 TTCTAAGGCTGTGAGACGAGGAGGAGCTGAGGCTGACAGACTGTGACATCAAC 1200  
 DB 1141 TTCTAAGGCTGTGAGACGAGGAGGAGCTGAGGCTGACAGACTGTGACATCAAC 1200  
 QY 1201 CACCCATGACATGACACCAATCAAGTAACTGAGAGCCCGTGAATACGCTGAT 1260  
 DB 1201 CACCCATGACATGACACCAATCAAGTAACTGAGAGCCCGTGAATACGCTGAT 1260  
 QY 1261 CAGGATTTGTGCTGACAGCTCCGATGATGTTTCCAACTGCTATATAGTCAAC 1320  
 DB 1261 CAGGATTTGTGCTGACAGCTCCGATGATGTTTCCAACTGCTATATAGTCAAC 1320  
 QY 1321 GACCATGAGGTGTGCTGACAGCTCCGATGATGTTTCCAACTGCTATATAGTCAAC 1380  
 DB 1321 GACCATGAGGTGTGCTGACAGCTCCGATGATGTTTCCAACTGCTATATAGTCAAC 1380  
 QY 1381 AAGATCCGAGAGAGCTGAGAGAGCAGTGTGCTGCTCCCGGACAGTCCCT 1440  
 DB 1381 AAGATCCGAGAGAGCTGAGAGAGCAGTGTGCTGCTCCCGGACAGTCCCT 1440  
 QY 1441 CCCACCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500  
 DB 1441 CCCACCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500  
 QY 1501 AGTGAAGTTCAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1560  
 DB 1501 AGTGAAGTTCAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1560  
 QY 1561 GAGCAGCTCAAAAGCCGTGACAGAGAGCTGAGCCCTCTCAGCCCGAGCAACAA 1620  
 DB 1561 GAGCAGCTCAAAAGCCGTGACAGAGAGCTGAGCCCTCTCAGCCCGAGCAACAA 1620  
 QY 1621 CCAAAAGGAG 1680  
 DB 1621 CCAAAAGGAG 1680  
 QY 1681 GAAAGTGAAG 1740  
 DB 1681 GAAAGTGAAG 1740

QY 1741 AATATACGACCAACGATATGTAGACAGAGACCAGCCCGCCATGAGACGACGCC 1800  
DB 1741 AATATATGACACCAACAGAAATGTAGAGAGAGAGCCAGCCCGCCATGAGACGACGCC 1800  
QY 1801 CCTCCACGATGTAGTGGAGAGAGAGACAAAGTCAAGCTATGTCTATGAGAGAGAG 1860  
DB 1801 CCTCCACGATGTAGTGGAGAGAGAGACAAAGTCAAGCTATGTCTATGAGAGAGAG 1860  
QY 1861 CGGACGCTCAGCTTGGACATCAACAGCTCCCGGCGAGAGCTGGGCGCGTGGTGCAC 1920  
DB 1861 CGGACGCTCAGCTTGGACATCAACAGCTCCCGGCGAGAGCTGGGCGCGTGGTGCAC 1920  
QY 1921 ATCATCCAGTCAGCGAGACCCCTCCCTGAGAAATTCACACCCGACGAGATTGAATGCAC 1980  
DB 1921 ATCATCCAGTCAGCGAGACCCCTCCCTGAGAAATTCACACCCGACGAGATTGAATGCAC 1980  
QY 1981 TTTTACAGCCCTGAGACCCGCTCCACACCTGCTGAGCTTGGAGCGCTATGTACACCTCTGTTT 2040  
DB 1981 TTTTACAGCCCTGAGACCCGCTCCACACCTGCTGAGCTTGGAGCGCTATGTACACCTCTGTTT 2040  
QY 2041 GCGGAGAGAGAGAGAGACCTTCAAGCTGAGAAATGTATGTATGCGCGGTTCTCCAA 2100  
DB 2041 GCGGAGAGAGAGAGAGACCTTCAAGCTGAGAAATGTATGTATGCGCGGTTCTCCAA 2100  
QY 2101 NATGAAAGGNTTCTCGGCTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160  
DB 2101 NATGAAAGGNTTCTCGGCTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160  
QY 2161 TTTTACAGAGGAG 2220  
DB 2161 TTTTACAGAGGAG 2220  
QY 2221 AAACCCGAG 2280  
DB 2221 AAACCCGAG 2280  
QY 2281 CCTTTTCCCCAG 2329  
DB 2281 CCTTTTCCCCAG 2329

RESULT 2  
AAS03029 standard; cDNA; 667 BP.  
XX  
AC AAS03029;  
XX  
DT 29-AUG-2001 (first entry)  
XX  
DE Human diagnostic and therapeutic (dlthp) cDNA sequence #18.  
XX  
KW Human diagnostic and therapeutic molecule; dlthp; gene therapy;  
KW thalassemia; cardiovascular disorder; cell proliferative disorder;  
KW cancer; neurodegenerative disorder; autoimmune disorder;  
KW infectious disorder; inflammatory disorder; developmental disorder;  
KW Incyte ID number 2432679dec; intracellular signalling molecule; ss.  
XX  
OS Homo sapiens.  
XX  
PN M0200121836-A2.  
XX  
PD 29-MAR-2001.  
XX  
PF 19-SEP-2000; 2000MO-US25643.  
XX  
PR 23-SEP-1999; 99US-0155760.  
PR 24-SEP-1999; 99US-0155939.  
PR 24-SEP-1999; 99US-0156294.  
PR 28-SEP-1999; 99US-0156565.  
PR 28-SEP-1999; 99US-0156624.  
PR 28-SEP-1999; 99US-0156625.  
PR 24-NOV-1999; 99US-0167410.  
PR 24-NOV-1999; 99US-0167453.

PR 24-NOV-1999; 99US-0167517.  
PR 24-NOV-1999; 99US-0167520.  
PR 24-NOV-1999; 99US-0167542.  
PR 29-NOV-1999; 99US-0167943.  
PR 29-NOV-1999; 99US-0167945.  
PR 30-NOV-1999; 99US-0168197.  
PR 30-NOV-1999; 99US-0168265.  
PR 30-NOV-1999; 99US-0168429.  
PR 30-NOV-1999; 99US-0168432.  
PR 01-DEC-1999; 99US-0168468.  
PR 01-DEC-1999; 99US-0168599.  
XX  
XX (INCYTE GENOMICS INC.  
PA Hodgson DM, Lincoln SE, Russo FD, Spiro PA, Banyille SC;  
PI Bratcher SR, Dufour GE, Cohen HJ, Rosen BH, Shah P, Chalup MS;  
PI Hillman JL, Jones AL, Yu YJ, Greenawalt LB, Panzer SR;  
PI Roseberry AM, Wright RJ, Chen W, Liu TF, Yap PE, Stockdreher TK;  
PI Amshay S, Fong WT;  
XX  
DR WPI; 2001-281607/29.  
XX  
XX Novel diagnostic and therapeutic polynucleotides, used in disease  
PT diagnosis and for gene therapy of conditions such as cancer and  
PT thalassemia  
PT  
PS Claim 1; Page 263; 299pp; English.  
XX  
XX The present sequence for human diagnostic and therapeutic (dlthp) cDNA  
CC sequence #18 is 1 of 71 (AAS03012-AAS03082) novel sequences described  
CC in the invention. The present sequence (Incyte ID No: 2432679dec)  
CC encodes an intracellular signalling molecule. The dlthp polynucleotides  
CC may be used to diagnose a condition disease or disorder associated with  
CC human molecules. They can be used to identify the presence of similar  
CC nucleic acids. Dithp polynucleotides may be used to generate hybridisation  
CC probes for use in chromosomal mapping. Polypeptides (Dlthp) encoded by  
CC dlthp are used to screen for molecules which bind to them and modulate  
CC their activity. Dithp polynucleotides can be used for gene therapy of  
CC disorders such as severe combined immunodeficiency syndrome (SCID),  
CC cystic fibrosis, thalassemia, haemophilia resulting from Factor VIII  
CC or IX deficiencies, cardiovascular disorders e.g familial  
CC hypercholesterolaemia (FH), cell proliferative disorders e.g. cancers,  
CC neurodegenerative disorders, autoimmune/inflammatory disorders,  
CC infectious disorders and developmental disorders. The antibodies can be  
CC used to analyse protein expression levels.  
XX  
SQ Sequence 667 BP; 229 A; 157 C; 149 G; 132 T; 0 other;  
Query Match 24.0%; Score 559.4; DB 22; Length 667;  
Best Local Similarity 98.8%; Pred. No. 1.8e-115;  
Matches 574; Conservative 0; Mismatches 6; Indels 1; Gaps 1;  
QY 4 GGGATCACTACATGTCTGCGGAGAGGCGCTT-GGGACGAGATTGAGAAATTCGCCAGT 62  
DB 10 GGGATCACTACATGTCTGCGGAGAGGCGCTTGGGCGGAGATTGAGAAATTCGCCAGT 69  
QY 63 AATGGGGATGAGCTAGAGAACTTCCCAATGTCTACACAGAGGCCAGGCCCAACCCCA 122  
DB 70 AATGGGGATGAGCTAGAGAACTTCCCAATGTCTACACAGAGGCCAGGCCCAACCCCA 129  
QY 123 GCCAGCCAAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 182  
DB 130 GCCAGCCAAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 189  
QY 183 CAAAGAGCAGACCAACCACTGCAATACCTCTCAGAGTGTGCTCAAGACACTATGGA 242  
DB 190 CAAAGAGCAGACCAACCACTGCAATACCTCTCAGAGTGTGCTCAAGACACTATGGA 249  
QY 243 ACACAGCTTTGATGCGCTTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 302  
DB 250 ACACAGCTTTGATGCGCTTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 309  
QY 303 TTAATAAGATCATTAAG 362

Dd	310	TTACATTAATCATCTTATAAAACGGCTATGATATATGGAAACATATAAGAACGGCTTGGAAAA	369
Oy	363	CAACATATTATCGAATAGCTCAGAAATGATATCCAGACCTTCAACACTATGTTTACAAATTG	422
Dd	370	CAACTATTACTGGAATGCTCAGAAATGATATCCAGACCTTCAACACTATGTTTACAAATTG	429
Oy	423	TTACATTTACAAACAGCTTGAGATGACATAGCTTTAATATGCAAGAGCTTGGAAAACCT	482
Dd	430	TTACATCTTACAAACAGCTTGAGATGACATAGCTTTAATATGCAAGAGCTTGGAAAACCT	489
Oy	483	CTTCTTGCAAAAAATTAATATGACTATCCACAGACGAAACCGATCATGATATGTCAGGC	542
Dd	490	CTTCTTGCAAAAAATTAATATGACTATCCACAGACGAAACCGATCATGATATGTCAGGC	549
Oy	543	AAAAAGAGAGGACGTGGAGAGAAAGAAACAGGACAGCAA	583
Dd	550	AAAAAGAGAGGACGTGGAGAGAAAGAAAGGAGGATATCAA	590
RESULT 3			
ABK83660	ID	ABK83660 standard; cDNA; 4664 BP.	
ABK83660:	AC		
14-AUG-2002 (first entry)	DT		
Human cDNA differentially expressed in granulocytic cells #231.	XX		
Human ss: granulocytic cell; DNA chip: bacterial infection;	KM		
fungal infection; parasitic infection; protozoal infection;	KM		
renal infection; sterile inflammatory disease; psoriasis;	KM		
rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;	KM		
cardiac reperfusion injury; renal reperfusion injury; ARDS;	KM		
adult respiratory distress syndrome; inflammatory bowel disease;	KM		
Croh's disease; ulcerative colitis; peridontal disease;	KM		
granulocyte activation; chronic inflammation; allergy.	XX		
Homo sapiens.	OS		
WO200228999-A2.	PN		
11-APR-2002.	PD		
03-OCT-2001: 2001WO-US30821.	PF		
03-OCT-2000: 2000US-237189P.	PR		
(GENE-) GENE LOGIC INC.	PA		
Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;	PI		
WPI; 2002-435328/46.	DR		
Detecting granulocyte activation by detecting differential expression	PT		
of genes associated with granulocyte activation, which serves as	PT		
diagnostic markers that is useful for monitoring disease states and	PT		
drug toxicity	PT		
Claim 1; SEQ ID No 231; 114pp; English.	PS		
The invention relates to detecting (M1) granulocyte (GC) activation	CC		
(GCA), by detecting the level of expression of gene(s) (Gs) identified by	CC		
DNA chip analysis as given in the specification, and comparing	CC		
the expression level to an expression level in an unactivated	CC		
GC, where differential expression of Gs is indicative of GCA.	CC		
Also included are modulating (M2) GA by contacting GC with an agent	CC		
that alters the expression of at least one gene in Gs; (2) screening (M3)	CC		
for an agent capable of modulating GCA or an inflammation (especially	CC		
chronic) in a tissue, an allergic response in a subject, exposure of a	CC		
subject to a pathogen or sterile inflammatory disease using the	CC		
gene expression profile; (3) detecting (M4) an inflammation (especially	CC		

chronic in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease, by detecting the level of expression in a sample of the tissue of gene(s) from GS, where the level of expression of the gene is indicative of inflammation; (4) treating (M5) an inflammation (especially chronic) or in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease, by contacting a tissue having inflammation with an agent that modulates the expression of gene(s) from GS in the tissue. M1 is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful for screening an agent capable of modulating GCA preferably in an inflammation in a tissue; M4 is useful for detecting an inflammation (especially chronic) in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease (e.g. psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal reperfusion injury, ARDS, adult respiratory distress syndrome, inflammatory bowel disease, Crohn's disease, ulcerative colitis, periodontal disease; also bacterial infection, viral infection, parasitic infection, protozoal infection, fungal infection and M5 is useful for treating one of the above conditions. The present sequence represents a gene differentially expressed in granulocytes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

Query Match	19.8%	Score 460.2	DB 24	Length 4664
Best Local Similarity	57.0%	Pred. No. 5.9e-93		
Matches 1151	Conservative 0	Mismatches 693	Indels 175	Gaps 10
QY 143	CCAAACCCCCCCCCCAGAGACCTTCCAACCCCTTAACAGCCCAAGAGGAGGAGACCAACCAAC	202		
Db 1877	CCAAACCCACCAACCCCCGGAGGTGTCCAAATCCCAAAAAGCAGAGGAGGTATCCAAACCAAGC	1936		
QY 203	TGCATATACCTTGTCTAGAGATGGTGTCTCAAGACACTATGTGAAACACCAATTTGGATGGCCTT	262		
Db 1937	TGCATATACCTTCAACACAGGTAGTGAAGAGGCTCTGTGGAACATCATGTTGGATGGCCAT	1996		
QY 263	TCCAGCACCTCTGTGATGCCGTCAAGCTGAACCTTCCCTGATTTACTATTAAGATCATTTAAAA	322		
Db 1997	TCCGGCACCTGTGGATCTGTCAAACTGGGGCTACCGGATATATCAAAAATTTATTAANAAC	2056		
QY 323	CGCCTATGGATATGTGGAAACAATAAAGAACGCTTGGAAAAACACTATTTACTGGAATGCTC	382		
Db 2057	AGCCTATGGACATGGGTACTTTAAAGAGAGACTTGA AAAACAATTTATTTGGCGTCTT	2116		
QY 383	AGGAATGATCCAGAGACTCAACACTATGTATTACAAATTTGTATCATCTACACCAAGAGCTG	442		
Db 2117	CAGAGTGTATCCAGATTTTAAATNCCATGTTCACCAACTGTTCATTTTAAACAACAGGCCCA	2176		
QY 443	GAGATGACATAGTCTTAAATGGCAGAGCTCTTGAAAAGCTCTTTCGCAAAAATAATTAATG	502		
Db 2177	CTGAATGATATTTGTCTTAATGGCACAACAGCTGGAAAAATATTTCTTACAGAAAGTTGAT	2236		
QY 503	AGCTACCCACAGAAACAAACCGAGATCATGATGTCCAGGCAAAAAGAAAGAGGAGAGCTGGGA	562		
Db 2237	CAATGCCACAAAGAAACAAAGAGCTGTGTAGTACCATCTCCCTTAAGAACACGCCCAAGAAAG	2296		
QY 563	GGAAGAAACAGAGGACAGCAAAAACCTGGACGTTTCCAGGATACCAACAAACAACTCAAGCAT	622		
Db 2297	GGGGCCAAATTTGGCAGCGCTCCAGGGCAGTGTATACAGATGCCATCAGGTGCTTGGCGTCT	2356		
QY 623	CGACTCCCTCCGCGACAGACCCCTCAAGCGGAATCTCTCTCTGTGTGCAAGGCGACGCTC	682		
Db 2357	-----CTTCTGTGTGCAACACAGACGCGTATATCTCTCCACCTGAAGTATACCTACCACTG	2410		
QY 683	ACCCCTTCCGCGCTCACCCCGGAGCTTCATCTGTCAAGACCCCTGTCAATGACAGTGTGTC	742		
Db 2411	TCTCTCAACATTTCCCAACCCATCATGCTATTTTC-----T 2443			
QY 743	CTCCCCAGCCACTGCAAGGCCGCCGCCAGTGGCCCCCAGCAGCACACCCCAACCCGCTC	802		



DB 2444 CTCCACCTTCTCAAGTCTTGTGACGTCGTGGAGACCCCGCTCTGTTGTTACTGCTGACGTC 2503  
QY 803 CAGCTCCCGCCCGTACAGAGCCACCACCATCATGCGGGCCACCACCCAGACGCTGTGA 862  
DB 2504 CTCGAGCCGACGCCCTTGTGCCAAGAAAAAGGCGT----- 2537  
QY 863 AGACAAAAGAGGAGTGAAGAGGAAGCAGACACCACCACCCACCAACCATTTAGACCCCA 922  
DB 2538 ----AAAGCGGAAAGAGATATCTACCAACCCCTACACCTACAGACATCTTGGCTCCTGTT 2593  
QY 923 TTACAGAGCCACCTCCCTGCGCCCGGAGAGCCAGACACCAACCTGGGCGC---AGCGGC 979  
DB 2594 CTCGAGTACGCCCTCGCGGGAGCTTTCAGGACGACGCGGCTTCCCTATGCGCTA 2653  
QY 980 GGGAGAGCAGCGGCTGTGTAAACCTCCAAAGAGAGCTCCGCGCATCTCAGAGACAC 1039  
DB 2654 GAGAGAGTGGTGGCCCATCAAGCCGCCACCAAGACTTGCCTGACTCTCAGCAACAC 2713  
QY 1040 CAGCAGCAGAGAGAGAGCAGAGAGGTCTCGAGAGCAGCTCAAGTGTGCGAGCGCATCTCA 1099  
DB 2714 ACCAGAGCTCTAAAGAAAGAAAGGCTTTCAGAAAGTAAACATTTGCAATGTGCAATTTGA 2773  
QY 1100 AGGAGATGTTGGCAAGAGACGCCCTTACGCGCTGCGCTTCTACAGCCTGTGGACG 1159  
DB 2774 AGGAGTACTCTTAAAGAGATGCTCTGCTATGCTTGTGGCTTCTTATTAACACAGTGTG 2833  
QY 1160 TGGAGCAGTGGGCTTACAGCAGTACTGTGACATCATCAAGCACCACCATGGACATGAGCA 1219  
DB 2834 CTCTGTGACATTTGGCTGCTGATGATGACATGATGATGATGATGATGATGATGATGATG 2893  
QY 1220 CAATCAAGCTAACTGAGAGCGCGGTGATGATGATGATGATGATGATGATGATGATGATG 1279  
DB 2894 CTGTCAAGCGGAAATGAGAGACGATGATGATGATGATGATGATGATGATGATGATGATG 2953  
QY 1280 TTGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1339  
DB 2954 TACGGCTTATGCTTCTCAAGCTGATGATGATGATGATGATGATGATGATGATGATGATG 1403  
QY 1340 TGGCCCGCAGCTTCCAGAGTGTGTGCAAAATGCGTTTGGCCAGATGCGGAGCAGAGCTG 1399  
DB 3014 TGGGACGAAGAGCTACAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 3073  
QY 1400 AGGAGCAGTGTGTGCGCGCTGCTCCCGGCGAGTGCCTCCCTCC---ACCAAGTGTGTG 1456  
DB 3074 TAGAACCAAGGCGCTTACAGCTCTCTACGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3133  
QY 1457 CCCCCGCTCATCCAGAGCAGAGCAGCAGTACTGCTCTC----- 1496  
DB 3134 CAGAGTCTCTCAGTGAAGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3193  
QY 1497 -----GGACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1540  
DB 3194 ATGAG 3253  
QY 1541 AGCGGCTGCTGAGCTCTCAG 1600  
DB 3254 ATGCTTGAAG 3313  
QY 1601 CTGAGCCCCGAGCAACAAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1659  
DB 3314 CCGGAGGTCTCAATATTCAG 3373  
QY 1660 GAAAAGCAAAAAG 1705  
DB 3374 GGAAGGAG 3433  
QY 1706 AAGCCAAAGAACTCTCTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1749  
DB 3434 CACCCCGCCCACTCAACCTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3493  
QY 1750 -----AGCAACAGCAATGTGAGCAAGA 1771

DB 3494 CTGCTTAAAGCCCTTCTGCGCTTGTGACCTTCTGAGAGAGTGGCAACCAAGTCCCAAAA 3553  
QY 1772 AGAGAGCAGCGCCCATGAAG 1831  
DB 3554 AGGCGACAAACAG 3613  
QY 1832 AGTCAAGCCTTATGCTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1891  
DB 3614 AGAGCAGGCGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3673  
QY 1892 CCGGCGAGAGAGTGGCGCGCTGTGTGATCATCATCATCATCATCATCATCATCATCATCAT 1951  
DB 3674 CTGGGAGAGAGCTGGCGCGAGTGTGATTAATTCAGAGAGAGAGAGAGAGAGAGAGAG 3733  
QY 1952 ATTCAACCCCGAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2011  
DB 3734 ATTCAACCCCGAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3793  
QY 2012 AGCTTGGAGCGCTATGTCACCTCTGTTTGGGAGAGAAA 2050  
DB 3794 AGCTT-GAGCGCTATGCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3831

RESULT 4  
ABN95159  
ID ABN95159 standard; DNA; 4664 BP.  
XX  
AC ABN95159;  
XX  
DT 13-AUG-2002 (first entry)  
XX  
DE Gene #1657 used to diagnose liver cancer.  
XX  
KW Gene: liver cancer; ds: hepatocellular carcinoma; hepatotropic;  
KW metastatic liver tumour; cytostatic; expression profile; disease state;  
KW disease progression; drug toxicity; drug efficacy; drug metabolism.  
XX  
OS Homo sapiens.  
XX  
XX WO200229103-A2.  
XX  
PN 11-Apr-2002.  
PD  
PF 02-Oct-2001; 2001WO-US030589.  
XX  
PR 02-Oct-2000; 2000US-237054P.  
XX  
PA (GENE-) GENE LOGIC INC.  
XX  
PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;  
XX  
XX WPI; 2002-426119/45.  
XX  
DR The invention relates to a novel method for diagnosing and detecting the  
XX progression of liver cancer, hepatocellular carcinoma or metastatic liver  
XX tumour in a patient, and differentiating metastatic liver cancer from  
XX hepatocellular carcinoma in a patient, involving detecting the level of  
XX expression of two or more genes represented in ABN95159-ABN97455 in a  
XX tissue sample. The method of the invention has hepatotropic, and  
XX cytoskeletal activity. The method is useful for diagnosing and detecting  
XX the progression of liver cancer, hepatocellular carcinoma and metastatic  
XX liver carcinoma in a patient. The method is useful for identifying  
XX expression profiles which serve as useful diagnostic markers as well as  
XX markers that can be used to monitor disease states, disease progression,  
XX drug toxicity, drug efficacy and drug metabolism.  
XX Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO  
CC at Itp.wipo.int/pub/published\_pcl\_sequences.

XX Sequence 4664 BP; 1084 A; 1336 C; 1293 G; 951 T; 0 other;

Query Match 19.8%; Score 460.2; DB 24; Length 4664;  
Best Local Similarity 57.0%; Pred. No. 5.9e-93;  
Matches 1151; Conservative 0; Mismatches 693; Indels 175; Gaps 10;

143 CCAACCCCGCCCGCCGAGACCTCCACCCCTAACAAGCCCAAGAGCGACAGCCACAC 202  
1877 CCAACCCCGCCCGCCGAGAGTGTCCATGCCCAAAAAGAGAGCGATTACCAACAGC 1936  
203 TCGAATCTGCTCAGAGTGTCTCAAGACATATGGAACCGAGTTTGCATGGCCTT 262  
1937 TCGAATCTGCTCAGAGTGTCTCAAGACATATGGAACCGAGTTTGCATGGCCTT 1996  
263 TCGAATCTGCTGATGTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 322  
1997 TCGGAGAGCTGTGATGTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 2056  
323 CCGCTATGATATGAGAACATTAAGAGCGCTTGGAAAAACAATTAATGATGCTGCT 382  
2057 AGCTATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2116  
383 AGAATGTATCCAGACTTCAACATGTTTCAAAATGTTTCAATCTCAACACAGCCTG 442  
2117 CAGAGGTATGCAAGATTTTAAATACATGTTCACCAATGTTTCAATTTCAACACAGC 2176  
443 GAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 502  
2177 CAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2236  
503 AGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 562  
2237 CAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2296  
563 GGAAGAAACAGGAGACAGCAAAACCTGGGTTTCCAGGTTTCAACACACATCTCAAG 622  
2297 GGGGCAAGTTGGAGGCTCCAGAGGAGTGTACAGAGTGTACAGAGTGTACAGAGTGT 2356  
623 CAGCTGCTCCGAGAGACCCAGACCCCTCAGACCGAATCTCTGCTGAGAGCCAGCCTC 682  
2357 CTTCTGTGTACACAGAGCCCTGTATATCTCTCCTCAGTGAATATCTACACTG 2410  
683 ACCCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 742  
2411 TCTCTAAATCTCCACCATCATGCTATTTCC-----T 2443  
743 CTCTCCAGCACTGAGAGCCCGCCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 802  
2444 CTCTCAATTTCTCAAGCTCTTGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2503  
803 CAGCTTCCAGGCGCTGACAGAGCCACCCATCATGCGGGCCACCCACAGCTGTGTA 862  
2504 CTCTCAGCCAGGCTCTTGGCAGAAAAAGGCT----- 2537  
863 AGACAAAGAGAGAGTGAAG 922  
2538 AAAAGGAG 2593  
923 TTTACAG 979  
2594 CTCTCAG 2653  
980 GGGAG 1039  
2654 GAG 2713  
1040 CAG 1099  
2714 ACCAGAGCTTCAAG 2773

1100 AGAGATGTTTGGCAAGAAAGACGCCCTAGCGCTGGCCCTTTCAAAAGCTGTGACG 1159  
2774 AGAGATGTTTGGCAAGAAAGACGCCCTAGCGCTGGCCCTTTCAAAAGCTGTGACG 2833  
1160 TGGAGCAGTGGGCTTACAG 1219  
2834 CTCTCAG 2893  
1220 CAATCAAGTCTTAACTGAG 1279  
2894 CTCTCAG 2953  
1280 TCGATGATGTTTCTCAACCTGATTAAGTACAACTCTCTGACCTGAGAGAGAGAGAGAG 1339  
2954 TACGCTTATGTTCTCAACTGCTATTAAGTACAACTCTCTGAGAGAGAGAGAGAGAG 3013  
1340 TGGCCGCAAGCTTCAAG 1399  
3014 TGGCAG 3073  
1400 AG 1456  
3074 TGAAGCAG 3133  
1457 CCGCCCTCTCATCAGCAG 1496  
3134 CAGAGTCTCTCAGTGAAG 3193  
1497 GAGAGTCTCTCAGTGAAG 3253  
3194 ATGAG 3253  
1541 AGCGCTGCTGAGCTGAG 1600  
3254 ATGCTTACAG 3313  
1601 CTGAGCCCAAG 1659  
3314 CCGAGGCTCCATATATCAG 3373  
1660 GAAAGCAG 1705  
3374 GGAAGGAG 3433  
1706 AAGCAG 1749  
3434 CAGCCGCGCAGCTCAACCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3493  
1750 -----AGCAACAGCAATGTGAGCAAG 1771  
3494 CTGCTTTAGGCTTCTGCTGCTTGTGAGCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 3553  
1772 AG 1831  
3554 AGGCAACAG 3613  
1832 AGTCAAGCTATGCTTATGAG 1891  
3614 AG 3673  
1892 CCGGAG 1951  
3674 CTGGGAG 3733  
1952 ATTCAAGCCGAG 2011  
3734 ATTCAAGCCGAG 3793  
2012 AGCTTGAAG 2050  
3794 AGCTT-GAGGCTATGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3831



Db	1488	ACGGCTTATGTTCTCCAACTGCTATAAGTCAATCCCGACGATCAGATGTGTGGCAAT	1547
OY	1341	GGCCCGAAGCTCCAGATGTGTTCCAAATGCCCTTTGGCAAGATGCGGAGACCCGTA	1400
Db	1548	GGCAGCAAAAGCTACAGGATGTATTGTAGTTCGGTTATGCCAAGATGCCAGATGAACCACT	1607
OY	1401	GGAGCCAGTGTGGCCGTGTCTCTCCCGGCAGTGCCTCTCCC---ACCAAGTTGTGGC	1457
Db	1608	AGAACACAGGGCTTTTACCAGTCTCTACTGCGCATGCCCCCTGGCTTGGCCAAATCGCTTTC	1667
OY	1458	CCCGCCCTTCATCCAGCCAGCAGCAGCAGCATAGCTCTCTCGACAGTGCAGTTGCATCGTA	1517
Db	1668	AGAGTCTCTCCAGTATAGGAAAGTAGACAGTAGACATCTCTTGAGGAAGAGGAGGAGAGGA	1727
OY	1518	TGACTCTGAGGAGAGCAGCAACCCAGG	1544
Db	1728	TGAGGAGACGAGGAGGAGGAGGAGAG	1754

	Query Match	Similarity	12.6%	Score 294.2	DB 19	Length 3104
	Best Local	Similarity	53.1%	Pred. No. 7.3e-16		
	Matches 995	Conservative	0	Mismatches 673	Indels 207	Gaps
OY	145	AACCCCGGCCCCAGAGACTTCAACCCTTAACCAACCCCAAGGGGACCAACCAATCG	204			
Dd	142	AACCTTCCTCCACACAGAATAATATAAATTACTAGAAAAAATGGCGATGTGCAAATCAACTT	201			
OY	205	CAATACCTGCTCAGAGTGTTGCTCAAAGACATATGGAAAACACAGTTTGATGGCTTTTC	264			
Dd	202	CAGTATCTCAAAAAAGTTGCTCCATAAGGAATTTATGGAACATAGTTTTCATGGGCCCTT	261			
OY	265	CACACACCTGTGGATCCGTCGAAGCGAAGCCCGCATTTACTATATAAGTTCATTAAAGC	324			
Dd	262	CACGCTCCGTGGATGCTGTGAAGAACTTAAGTTGCCGATTAATTAACAATTATAAAAC	321			
OY	325	CCATGATATGGGAACATATAAAGAGCGCTTGAAGAAACACTATATACGAAATCG	384			
Dd	322	CCAATGGATTTAAATCAATTAAGAAAGCGCTTGGAAATATAATATATGCGAAGCTTCA	381			
OY	385	GAATGATCCAGAGACTTCAACACTATGTTTACAAATTTGTTACATCTACAAACAGCTTGA	444			
Dd	382	GATGATATGAAGACTTCAATCAATATGTTCTCAAAATTTGTTATTTATTAACAAAGCTTGA	441			
OY	445	GATGACATAGTCTTATGGCAGAGCTCTGGAAAAAGCTTTTCCAAAAATAAATGAG	504			
Dd	442	GATGATATTTGTTCTTATGGCACACACTCTTAGAAGAGCTGTTATTCAGAAATATCTGAG	501			
OY	505	CTACCCACAGAAANAACCGAGATCATGATATGTCAGGCGAAAAGAAAGAGAGCTGGAGG	564			
Dd	502	ATGCCACAAAGAG-----AGCAATTTGTGGTGTTAAGSAAAGA	540			
OY	565	AAAGAAACAGSGAGACAGCAAAACCTGGCGTTTCCAGCGTACCAAAACAATCTMACATCG	624			
Dd	541	ATCAAGAAAGGCACTCAACAGATATAGTGTTTCTTCTGCTAAAGAAATAATCATACACC	600			
OY	625	ACTGCTCCGCAACACCCAGACCCCTCAGCGGAATCTCTCTGCTGTCAGGCCACGCTTAC	684			
Dd	601	ACGCGCAAGAAAAGTATTTAAGGAGAGAAATTTCTTGATTTCTCAAGACA---	657			
OY	685	CCCTTCCTGCGCTACCCCGGACCTCATGTCAGACACCCCTGTATGACAGTGTGCT	744			
Dd	658	-----	657			
OY	745	CCCAGGACACTGAGAGCGCCCCCGCAGTGCCCCCCAACCAACCCCAACCGCTCCA	804			
Dd	658	-----TCTATT	663			
OY	805	GCTCCCGACCGCTTACAGAGCCACCACCATCATGCGGGCCACCCCAACGCTGTGAG	864			
Dd	664	TCTCTTTGAACGTGTGTACAGGAGCTTCAGTCACTCAATTCANAAATCGCGCCAA	723			
OY	865	ACAAAGAAGGAGTAAAGAGAAAGACAGACACACACCCACACCATTAAGCCCAT	924			
Dd	724	GTTTACAAAAGCTGTGAABAGAAAGACACATACACAACTCTTGC AAC--TTACAGAGTTA	781			
OY	925	CACGAGCAACCTCGCTCCCGCCGAGGCCAAGACACCAAGCTGGGCAACGGCGGAG	984			
Dd	782	AAGCAAGTAGTAATTTCTCCAAATTCACAGAAAAATAGTGGCACTGCACTATAA	841			
OY	985	AGCAGCGCGCTGTGAACCTTCCAAAGAGACGTECCCGACTCTCAGCAGCACCCAGCA	1044			
Dd	842	AAGAA-----AATATGCCAAAGAAATTTTTGTCAGATTCAGACACAAATATAT	891			
OY	1045	CCAGAGAGAGAGCAAGAGTCTCGSAGAGAGTCAAGTGTCTGACGGGATCTCTAAGAG	1104			
Dd	892	GTTGTGAGACTGTTAAAGTAACGTAAACAAATTAAGGCACTGTAGATTCCTTAAGAA	951			
OY	1105	ATGTTTGCAGAAAGACAGCGCGCTGAGCGCTTCTTCAAGAGCGTGGAGTGGAG	1164			
Dd	952	ATGCTTGAAGAAACATTTTTCATATGACATGGCCCTTTTATAATCTCTGTGAGCTTAAT	1011			
OY	1165	GCACTGGGCTACACGACTACTGTGACATCATCAAGACCCCATGAGACATGAGCACAATC	1224			

```
Db 1012 GCTTTGGACCTCCATPACTACTATGAGCTTGTCAAAAATCCAGAGATCTTGGAACTATT 1071
QY 1225 AAGTCTAACTGAGGCCCTGAGTACCGTATGCTAGAGTTGGTGTGAGCTGCCGA 1284
Db 1072 AAGGAGAAATATGATTAACCAAGATATATAGATGATCTATTTGGCGCAGATGTAGA 1131
QY 1285 TTGATGTTCTCAACTGCTATATAGTACAACTCTCTGACATGAGTGTGGCCATGGCC 1344
Db 1132 TTATGTTTCATGAAATGCTACAGTACAACTCTCTGACATGAGTGTGGCCATGGCA 1191
QY 1345 CGCAAGCTCCCGAGTGTCTTGAATGCGTTTCCCAAGATCCCGGAGACGACCTGAGAG 1404
Db 1192 AGAATGCTCAGATGTTTCTGAAACGATTTTCAAGATCCCGATTTGAACCTGTTAG 1251
QY 1405 CCAGTGTGCGCTGTGCTCCCGGAGTGGCCCTCCCAAGAGTGTGGCCCGCC 1464
Db 1252 AGTATGCTTTATTTATACATAAAGATATTCACAGAAACCACTGCT----- 1299
QY 1465 TCATCCAGCAGCAGCAGCAGCAGTACCTCTGAGCAGTACAGTTCAGTATGATCTCT 1524
Db 1300 -----AGAGAGAACACTAATGAAGCCCTCTGTGAAGGGAAC---TCTTGTGATGATCT 1350
QY 1525 GAGGAGAGAGCAGCCAGCCGCTGCTGAGCTCCAGAGAGCAGTCAAAAGCCGTGACAG 1584
Db 1351 GAAGATAGAGCAGATTAAGCCGTCTGCAAAAGCTTCAAGGAGCTTAAAGCTGTACATCA 1410
QY 1585 CAGCTTCAGCCCTCTCTCAGCCAGCAGCAAAACCAAGCAAAAGAGAGCAAAAGACAAG 1644
Db 1411 CAGCTCCAGGTTTGTGTCCCAAGTACCTTTCCTGAAGCTAAATTAAG----- 1458
QY 1645 AAGGAAAAAGAAAAAGCAACAAAAAGAAAGAGAGATGGAAGATTAAGAAAAAGC 1704
Db 1459 -----AAAGAAAGCTTAATAAAGAAAAAGAAAAAGAAAGGTTAATAAGAGC 1506
QY 1705 AAAGCAAGAGAACTCTCTCTTAAAAAGCAGAAAGAAATATAGCAGCAGCAATGTG 1764
Db 1507 AATATAAA-----TCCAAAGAAAAATGTGTAGCAGCAAAATGAGGCTAAAGAAAAAGTCC 1557
QY 1765 AGCAAGAAAGAGCAGCCAGCCATGAAGAGCAAGCCCTCCAGATGATGATCGAGAGAA 1824
Db 1558 AAGGAAATATCAGCAAAAGAAAGAAACAAACAGTTGATGTCTAAAA---TCTGAAGAT 1614
QY 1825 GAGGACAAAGTGCAGACCTATGTCTATGAGAGAGAGCGGAGCTAGCTTGACATCAAC 1884
Db 1615 GAAGATATATGCTTAACCTATGATGATGAGAAAAAGCGATTAAGCTGAATATATAAC 1674
QY 1885 AAGTCCCCCGGAGAGCTGCGGTGCTGACATCATCCAGTACCGGAGACCTCC 1944
Db 1675 AACTCTCCCTGAGATTAACCTTGGCGGAGTACTTCACTAATATACATCAAGAGACCTCT 1734
QY 1945 CTGAAGAAATTCACACCCGAGCAGATTAATGCAATTCGATTTGAGACCCGAGCCACA 2004
Db 1735 CTGAGCAATTCATCTGATGAGATGAGATAGATTAACCTTTGAACACTGAAAGCATCAAC 1794
QY 2005 CTGCGTAGCTTGA 2019
Db 1795 CTAGAGCAATTAGAA 1809

RESULT 7
AAK40172
ID AAK40172 standard; DNA: 3106 BP.
AC AAK40172:
XX 02-JUL-1999 (first entry)
XX DE MO9904265 Seq ID No: 686.
XX KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
```

```
XX OS Homo sapiens.
XX PN MO9904265-A2.
XX PD 28-JAN-1999.
XX PE 15-JUL-1998; 98WO-US14679.
XX PR 22-JUN-1998; 98US-0102322.
XX PR 17-JUL-1997; 97US-0896164.
XX PR 10-OCT-1997; 97US-0061599.
XX PR 10-OCT-1997; 97US-0061765.
XX PR 11-OCT-1997; 97US-0948705.
XX PR 11-OCT-1997; 97GB-0021697.
XX PA (LUDW-) LUDWIG INST CANCER RES.
XX PI Chen Y, Gout I, Gure A, O'Hare M, Ohta Y, Old LJ;
XX PI Pfeundschuh M, Sahlin U, Scanlan MJ, Stockert E;
XX PI Tureci O;
XX PT WPI: 1999-132448/11.
XX PS New isolated cancer associated nucleic acids and polypeptides -
XX PT isolated using sera from cancer patients, used to develop products
XX PR for the diagnosis, monitoring or treatment of cancers
XX PS Claim 67; Page 730-731; 787pp; English.
XX CC The invention relates to a method for diagnosing a disorder characterised
XX CC by expression of a human cancer associated antigen precursor coded for by
XX CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
XX CC biological sample isolated from a subject with an agent that specifically
XX CC binds to the NAM, an expression product or a fragment of an expression
XX CC product complexed with an HLA molecule; and (b) determining the
XX CC interaction between the agent and the NAM or the expression product as a
XX CC determination of the disorder. The products and methods can be used in
XX CC the diagnosis, monitoring, research, or treatment of conditions
XX CC characterised by the expression of various cancer associated antigens.
XX CC The invention provides nucleic acid sequences and encoded polypeptides
XX CC which are cancer associated antigen precursors expressed in human breast
XX CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
XX CC lung cancer.
XX SQ Sequence 3106 BP; 1179 A; 560 C; 594 G; 773 T; 0 other;

Query Match 12.6%; Score 294.2; DB 20; Length 3106;
Best Local Similarity 53.1%; Pred. No. 7.3e-56;
Matches 995; Conservative 0; Mismatches 673; Indels 207; Gaps 9;

QY 145 AACCCCGCCCGCCAGAGACCTCCAAACCTTAACGCCCAAGAGCAGACCAACCACTG 204
Db 144 AACCTCTCCACCAAGATATATATTAATTAAGAAAAATGGCGCATTAACAAATCACTT 203
QY 205 CAATACCTGCTCAGAGTGTGCTCAAGCACTATGAGAAACACCAAGTTGATGGCTTTC 264
Db 204 CAGTATCTACAAAAAGTGTCTTAAGATTAAGATTAAGCAATAGTTTTCATGGCCCTTT 263
QY 265 CAGCAGCCTGTGATGCCGTCAAGCTGAACCTCCCTGATTACTATAGCAATTTAAAG 324
Db 264 CAAGCTCCTGTGATGTGTAAGAACTAAAGTTGCTGTATTATATACATTAATAAAG 323
QY 325 CCTATGATATGGAGCAATTAAGAGCGCTTGAAGAAACCACTATATAGCAATGCTCAG 384
Db 324 CCAATGATTTTAATATCAATTAAGAGCGCTTGAAGAAATTAATATAGCGAAGCTTCA 383
QY 385 GAATGTATCAGACTTCAACACTATGTTTACAAATTTGTACATCAACAAAGCTGGA 444
Db 384 GAATGTATAGAAAGCTTCAATCAATGTTTCTCAAAATTTTATTAATAACAGCTGGA 443
QY 445 GATGACATAGTCTTATATGAGCAAGAGCTGTGAAAAAGCTCTTTCGCAAAAAATAATAG 504
Db 445 GATGACATAGTCTTATATGAGCAAGAGCTGTGAAAAAGCTCTTTCGCAAAAAATAATAG 504
```

D 444 GATGACATTGTTCTTATGGACAAGCTCTAGAGAGCTGTTTATGCAGAAATTATCTCAG 503  
Q 505 CTACCCACAGAGAAACCGAGATCATGATAGTCCAGGCAAAAGAGAGAGAGCTGGAG 564  
D 504 ATCCACAGAGAG-----AGCAAGTTGGGTGTTAAGGAAGA 542  
Q 565 AAGAAAGAGGAGCAGCAAAACCTGGCGTTTCCACGGTACCAACACACTCAAGCTCG 624  
D 543 ATCAAAAGAGGACACTAACGAAATATAGCTGTTTCTTCTTAAGAAATATCATCC 602  
Q 625 ACTCTCCGAGAGACCCCTCAGCCGATCTCTCTCTGTCAGGCGACGCTCAC 684  
D 603 AGGCAACAGAAAAAGATTATTAAGCAGCAAGAAATCTCTCTATTTCTTAAGACA--- 659  
Q 685 CCCCTCCCTGCGCTCAGCCCGGACCTCATCTGTCAGAACCCCTGTCATGACAGTGTGCT 744  
D 660 ----- 659  
Q 745 CCCGACCACTGCAGAGCCCGCCGAGTCCCGCCCGCCAGCAACCCCAACCCGCTCCA 804  
D 660 -----TCTATT 665  
Q 805 GCTCCCAAGCCCTTACAGAGCACACCACCATCATCGCGGCAACCCACAGCCCTGGAAG 864  
D 666 TCTCCCTTGAACTGGTACAGGAGCTTCACTCACTCCAGTTCAACAACTGGGCCCCA 725  
Q 865 ACAAGAGAGGAGTGAAGAGAAAGACAGACACCAACCCACACCATTTGACCCCTT 924  
D 726 GTTACAAAGAGTGAAGAGAAAGAGATCAACAACCTCTGCAAC--TTGACAGTTA 783  
Q 925 CAGAGCCACCCCTGCGTGGCGCCGAGCCCAACACCAAGCTGGGCGAGCCGCGGAG 984  
D 784 AAGCAAGTATGTAATTTCTCCAACTTCAACAAAAATCAGTGGCCTCCACTATTA 843  
Q 985 AGCAGCCGCGCTGTGAACCTCCAAAGAGAGAGCTGCCACTCTGACAGCACCCAGCA 1044  
D 844 AAGAA-----AATATGCCAAAGAAATGTTTGGCAATTTCTCAGCAACAATATAT 893  
Q 1045 CCAGAGAGAGAGAGCAAGGCTCTGGAGAGCTCAAGTCTGACGCGGACTCTCAAGAG 1104  
D 894 GTTGTGAGAGCTGTTAAAGTACTGAACATTAAGGACGTGATGATTTTAAAGAA 953  
Q 1105 AATGTTGCCAAGAGAGACCGCGCTACGCGCTCTTCAAGCCCTGAGACCTGCGAG 1164  
D 954 ATGCTTGCAGAGAAACATTTTATATGATGAGGCGCTTATATCTCTGTTGACGTTAAT 1013  
Q 1165 GCACCTGGCTTACACAGACTGTGATCATCATGACAGACCCATGACATGACACATC 1224  
D 1014 GCTTGGAGACTCATACTACTCATGACGTTGTCAAAAATCCGATGATCTTGAACATAT 1073  
Q 1225 AAGCTTAACCTGAGGCGCGGTAGTACCGTATGCTCAGAGATTGGTGTGACGTCGA 1284  
D 1074 AAGGAGAAATATGATTAACGAAGATTAAGAGTACTATCATATGCGCAGATGTAGA 1133  
Q 1285 TTGATGTTCCCAAGTATATAGTACAAACCCCTGACACATGAGGCGCGGACCC 1344  
D 1134 TTATATGATATGATTGCTACAAAGTACAAATCTCCAGATCAGAAAGTTGGAATAAGCA 1193  
Q 1345 CGCAAGCTCCAGAGATGTTTGAAGATGCGCTTGCAGAGATGCGGAGAGAGCTGAGAG 1404  
D 1194 AGAATCTTCAGAGATGTTTTCGAAAGCAATTTTCAAAAGATCCGATGGAACCTGTTAG 1253  
Q 1405 CCAAGTGTGCGCTGTCTCCCGGAGTGCCTCCACCAAGGTTGTGGCCGCGCC 1464  
D 1254 AGTATGCTTTATGTTATCATCAAAACAGATATCAGAAACCACTGCT----- 1301  
Q 1465 TCATCCAGAGAGAGAGAGAGAGTACTCTCTGACAGAGTTCGACTGATGACTCT 1524  
D 1302 -----AGAGAGAACTATATGAGGCTCTCTGAAAGGGAAC---TCTTCTGATGATTC 1352  
Q 1525 GAG 1584  
D 1353 GAAGATGAG 1412

Q 1585 CAGCTTGCAGCCCTCTCTCAGGCCAGCAGACAAACCAAGAAAAAGAGAAAGCAAG 1644  
D 1413 CAGCTCAGAGTTTGTGCCAAGTACTCTTCCGTAAGCTAATATAAAG----- 1460  
Q 1645 AAGAAAAAGAAAAAGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1704  
D 1461 -----AAGAGAAAGTAAAGAAAAAGAAAAAGAAAAAGGTTAATTAACAGC 1508  
Q 1705 AAGCCAAAGAGCTCTCTCTTAAAGAGAGAGAAAAATTAATGACAGCAACAGCAATGTG 1764  
D 1509 AATGAAAA-----TCAAGAAAAATGTGTGAGCAATAGAGCTTAAGGAAAGTCC 1559  
Q 1765 AGCAAG 1824  
D 1560 AAGAGAAATACAGCCAAAGAAAAAGAAAAAGAAACAACTTATGCTGCTAATAA---TCTGAAGAT 1816  
Q 1825 GAGGACAGAGTGAAGCTTATGCTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1884  
D 1617 GAAGATTAATGCTTAACCTATGACACTATGATGAGAGAAAAAGCAGTTAAGTGAATATAAC 1866  
Q 1885 AAGTCCCGCGAG 1944  
D 1677 AAGCTCCCTGAGATTAACCTTGGGCGAGTATGCTTACATTAATACAAATCAGAGAGAGAG 1736  
Q 1945 CTGAAGATTTCAACCCCGAG 2004  
D 1737 CTGAGCAATTCATCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1796  
Q 2005 CTGCTGAGCTTGA 2019  
D 1797 CTAAGAGAAATTAAGAA 1811

RESULT 8  
ID ABV25740 standard; cDNA; 1862 BP.  
XX  
AC ABV25740;  
XX  
PF 16-SEP-2002 (first entry)  
XX  
DE Human prostate expression marker cDNA 25731.  
XX  
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
KW pharmacogenomic marker; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200160860-A2.  
XX  
PD 23-AUG-2001.  
XX  
PF 20-FEB-2001; 2001WO-US05171.  
XX  
PR 17-FEB-2000; 2000US-18319P.  
XX  
PR 16-MAR-2000; 2000US-189862P.  
XX  
PR 25-MAY-2000; 2000US-207454P.  
XX  
PR 09-JUN-2000; 2000US-211314P.  
XX  
PR 18-JUL-2000; 2000US-219007P.  
XX  
PR 13-DEC-2000; 2000US-255281P.  
XX  
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
PI Schlegel R, Endege WO, Monahan JE;  
XX  
DR WPI; 2001-662795/76.  
XX  
PT Novel isolated nucleic acid molecule associated with cancerous state of  
XX prostate cells and correlating with presence of prostate cancer, useful  
XX for detecting presence of prostate cancer, stage of prostate cancer -  
PS Claim 1; Page 5167-5168; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (I) is useful for:  
CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate  
CC cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
CC in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
CC patient;  
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.  
XX  
XX  
SQ Sequence 1862 BP; 502 A; 514 C; 483 G; 363 T; 0 other;

Query Match 9.8%; Score 227.8; DB 23; Length 1862;  
Best Local Similarity 58.6%; Pred. No. 4,4e-41;

Matches 555; Conservative 0; Mismatches 312; Indels 80; Gaps 6;

QY 841 GCGGCCACCCACAGCCTGTGAAGACAAGAGGAGTGAAGAGAAAGCAGACACACC 900  
DB 389 GCTCTCCACCCACAGCCTGTGAAGACAAGAGGAGTGAAGAGAAAGCAGATCTAC 448  
QY 901 ACCCCACACACATGACCCCAT-----TCACGAGCACCCTCGCTGCCCGG 948  
DB 449 ACCCTACACCTACAGCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 508  
QY 949 GAGCCCAAGACAAGAGCCTGAGGCGC--AGCGGGGAGAGAGCAGCAGCCTGTGAACT 1005  
DB 509 GAGCCTAAGAGCAGCAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 568  
QY 1006 CCAAGAGAGAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1065  
DB 569 CCAAGAGAGAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 628  
QY 1066 TCGAGAGAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1125  
DB 629 TCGAGAGAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 688  
QY 1126 GCTGAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1185  
DB 689 GCTGAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 748  
QY 1186 TGTGACATCATCAAGAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1230  
DB 749 CATGACATCATCAAGAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 808  
QY 1231 -----AACTGAGAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1280  
DB 809 TGTGAG 868  
QY 1281 CCGATGATGCTTCTCAAGTGTATTAAGTAAACCTGCTGACCATGAGTGTGCGCAT 1340  
DB 869 ACCGCTATGCTTCTCAAGTGTATTAAGTAAACCTGCTGACCATGAGTGTGCGCAT 928  
QY 1341 GGGCCGAAGCTCAGAGATGCTTGAAGAGCCTGCTGCTGCTGCTGCTGCTGCTGCT 1400  
DB 929 GGGACGAAG 988  
QY 1401 GGAAGCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1457  
DB 989 AGAAGCAG 1048  
QY 1458 CCGGCTCTCATCAAG 1496  
DB 1049 AGAGTCTCTCAAG 1108  
QY 1497 -----GGAGAGTGAAGTGTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1541

DB 1109 TGAGAGAGCAG 1168  
QY 1542 GCGGCTGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 1601  
DB 1169 TCGCTTACAG 1228  
QY 1602 TCAGCCCGAG 1660  
DB 1229 CCAGGCTTCATATTCACAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1288  
QY 1661 AAAAGCAG 1707  
DB 1289 GAAAGCAG 1335  
RESULT 9  
ID AAS26606  
XX AAS26606 standard; cDNA; 424 BP.  
AC AAS26606;  
XX  
DT 07-NOV-2001 (first entry)  
DE Human cDNA encoding a novel secreted protein, Seq ID 785.  
XX  
XX Human: immunosuppressive; antifibrillic; ss; antirheumatic;  
KW cytosolic; cardiac; vasotropic; cerebroprotective; nootropic;  
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;  
KW valvular; secreted protein; rheumatoid arthritis;  
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;  
KW cerebrovascular disorder; cerebral ischaemia; angiodysplasia;  
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;  
KW corneal infection; wound healing; epithelial cell proliferation;  
KW skin ageing; food additive; preservative; antiproliferative.  
XX  
OS Homo sapiens.  
XX  
PN W020015322-A2.  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01341.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225216.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.













```
RESULT 12
AA66021
ID AAF66021 standard; cDNA; 352 BP.
XX
AC AAF66021;
XX
DT 09-APR-2001 (first entry)
XX
DE Novel human polynucleotide, SEQ ID NO: 1777.
XX
KW Human; cytostatic; gene therapy; colon cancer; prostate cancer;
XX breast cancer; lung cancer; cancer detection; ss.
OS Homo sapiens.
XX
PN WO200102568-A2.
XX
PD 11-JAN-2001.
XX
PF 30-JUN-2000; 2000WO-US18374.
XX
PR 02-JUL-1999; 99US-0142310.
XX PR 02-JUL-1999; 99US-0142311.
XX
PA (CHIRON ) CHIRON CORP.
XX (HYSE-) HYSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;
PI Crkenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D;
PI Kita D, Garcia V, Jones LM, Strache-Crain B;
XX
DR WPI; 2001-091805/10.
XX
PT Library of polynucleotides for diagnosing a cancerous state of a
XX mammalian cell and detecting cancer, particularly of the colon or
XX prostate, comprises 3351 human polynucleotide sequences -
XX
PS Claim 9; Page 798; 1046bp; English.
XX
CC The present sequence is one of 3351 sequences in a library of human
XX polynucleotides. The library is used to detect differentially expressed
XX genes correlated with a cancerous state of a mammalian cell and can
XX detect colon, prostate, breast and lung cancer. The library can be used
XX to produce probes for detection of mRNA and to produce additional copies
XX of the polynucleotides. The probes can be used for chromosome mapping of
XX the polynucleotide and for detection of transcription levels. Ribozymes
XX or antisense oligonucleotides can be generated. The polynucleotides and
XX their gene products are used as genetic or biochemical markers (e.g. in
XX blood or tissues) that will detect the earliest changes along the
XX carcinogenesis pathway and/or monitor the efficacy of therapies and
XX preventive interventions. The polynucleotides, polypeptides and
XX antibodies against them can be used in pharmaceutical compositions to
XX treat the cancers and proliferative disorders such as neoplasia,
XX dysplasia and hyperplasia.
XX
SQ Sequence 352 BP; 94 A; 144 C; 83 G; 31 T; 0 other;
XX
Query Match 8.9%; Score 207; DB 22; Length 352;
Best Local Similarity 100.0%; Pred. No. 1.1e-36;
Matches 207; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 925 CACGAGCCACCTCGCTGCCCGGAG 951
DB 325 CACGAGCCACCTCGCTGCCCGGAG 351
DB
RESULT 13
ABL04337
ID ABL04337 standard; cDNA; 6776 BP.
XX
AC ABL04337;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 7493.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PMD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT P-PSDB; ABB60234.
XX
PS New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
CC Claim 1; SEQ ID NO 7493; 21bp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX
CC The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 6776 BP; 1749 A; 1878 C; 2060 G; 1089 T; 0 other;
XX
Query Match 8.8%; Score 203.8; DB 23; Length 6776;
Best Local Similarity 73.8%; Pred. No. 1.7e-35;
Matches 259; Conservative 0; Mismatches 92; Indels 0; Gaps 0;
```



```

Db 306 GCGTAGAGAGATGGTGGCCCCATCAAGCCCCCAAGCAAAAGACTTGGCTGACTCTCAGCA 247
QY 1035 GCACCCAGACACAGAGAGAGAGCAAGGTCTCGAGCAGCTCAAGTCTGACAGCGGCAAT 1094
Db 246 ACAACACACAGAGCTTTAAGAAAGAAAGCTTTCAGAACAGTTAAACATTGCAATGSCAT 187
QY 1095 CCTCAAGGAGATGTTGGCCAAAGAACGACGCGGCTACGCCCTGGCCCTTTCACAAGCCTGT 1154
Db 186 TTTGAAGAGAGTTACTCTCTAAGAGCAATGCTGCTATGCTTGGCCTTCTATTAACCAAT 127
QY 1155 GGAAGTGGAGCACTGGGCTTACACGACTGATGATCATCAAGCACCCTCAATGACAT 1214
Db 126 GGATGCTTTCGCACTGGCCTGCATGACATCAATGACATCATTAAGCACCCTCAATGACAT 67
QY 1215 GAGCACAATCAAGCTTAAGCTGAGAGGCGCGTGAATACCGTGAATGCTCAGAGTTTGGTGC 1274
Db 66 CAGCACTGTCAGCGGAAGATGAGAAACCGTGATTAACGGGATGACACAGAGTTTGGTGC 7
QY 1275 TGACGT 1280
Db 6 TGATGT 1

```

Search completed: March 1, 2003, 13:46:34  
 Job time : 506 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 1, 2003, 11:56:39 ; Search time 87 Seconds  
(without alignments)  
8209.772 Million cell updates/sec

Title: US-09-700-590A-101

Percent score: 2329

Sequence: 1 gcagggatcactagcatgctc.....ngccttaaaagtlnccca 2329

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents.NA.\*

1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCtus.COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfile1.seq.\*

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	159.2	6.8	450	US-09-328-111-516	Sequence 516, App
2	104.8	4.5	429	US-09-397-787-243	Sequence 243, App
3	84.6	3.6	7218	US-08-232-463-14	Sequence 14, Appl
4	64.8	2.8	2223	US-08-257-073-4	Sequence 4, Appl
5	64.8	2.8	8147	US-09-514-247A-9	Sequence 9, Appl
6	63.4	2.7	320	US-09-165-264-14	Sequence 14, Appl
7	63	2.7	320	US-09-165-264-7	Sequence 7, Appl
8	62.8	2.7	318	US-09-165-264-12	Sequence 12, Appl
9	62.6	2.7	320	US-09-165-264-11	Sequence 11, Appl
10	62.6	2.7	16442	US-08-781-891-208	Sequence 208, App
11	61.4	2.6	320	US-09-165-264-13	Sequence 13, Appl
12	60.6	2.6	319	US-09-165-264-8	Sequence 8, Appl
13	59.4	2.6	13987	US-08-804-227C-13	Sequence 13, Appl
14	59.4	2.6	44377	US-08-804-227C-7	Sequence 7, Appl
15	59.4	2.6	44377	US-08-804-198-1	Sequence 1, Appl
16	59	2.5	4403765	US-09-103-840A-2	Sequence 2, Appl
17	58.2	2.5	12980	US-08-811-566-5	Sequence 5, Appl
18	58.2	2.5	12980	US-09-034-756-5	Sequence 5, Appl
19	57.2	2.5	1798	US-09-797-906-1	Sequence 1, Appl
20	57	2.4	2447	US-09-014-969-14	Sequence 14, Appl
21	56.2	2.4	153331	US-09-128-185-16	Sequence 16, Appl
22	55.6	2.4	43795	US-08-742-185-101	Sequence 101, App
23	55	2.4	356	US-08-520-678A-22	Sequence 22, Appl
24	55	2.4	356	US-08-897-126-22	Sequence 22, Appl
25	55	2.4	9646	US-08-811-566-1	Sequence 1, Appl
26	55	2.4	9646	US-09-034-756-1	Sequence 1, Appl
27	54.6	2.3	240	US-08-628-417-6	Sequence 6, Appl

28	54.6	2.3	452	4	US-09-397-787-147	Sequence 147, App
29	54.2	2.3	7326	1	US-08-194-468-1	Sequence 1, Appl
30	54.2	2.3	7326	1	US-09-514-247A-7	Sequence 7, Appl
31	54.2	2.3	7344	3	US-08-961-739-1	Sequence 1, Appl
32	54.2	2.3	80246	4	US-09-078-294-4	Sequence 4, Appl
33	54.2	2.3	80595	4	US-09-078-294-3	Sequence 3, Appl
34	53.8	2.3	3275	4	US-09-370-838-151	Sequence 151, App
35	53.8	2.3	9589	1	US-07-925-695-1	Sequence 1, Appl
36	53.8	2.3	9589	1	US-07-925-695-2	Sequence 1, Appl
37	53.6	2.3	1051	4	US-09-245-041-10	Sequence 10, Appl
38	53	2.3	12001	1	US-08-458-568A-11	Sequence 11, Appl
39	52.8	2.3	1447	4	US-09-443-041A-27	Sequence 27, Appl
40	51.8	2.2	2674	4	US-09-817-180-1	Sequence 1, Appl
41	51.6	2.2	260	2	US-08-520-678A-29	Sequence 29, Appl
42	51.6	2.2	260	4	US-08-897-126-22	Sequence 22, Appl
43	51.6	2.2	980	4	US-09-171-209-8	Sequence 8, Appl
44	51.2	2.2	376	2	US-08-623-906A-18	Sequence 18, Appl
45	51.2	2.2	405	2	US-08-299-074A-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1  
US-09-328-111-516  
Sequence 516, Application US/09328111  
Patent No. 6262333

GENERAL INFORMATION:

APPLICANT: Endege, Wilson O.  
APPLICANT: Steinmann, Kathleen E.  
APPLICANT: Astle, Jon H.  
APPLICANT: Burgess, Christopher C.  
APPLICANT: Bushnell, Steven E.  
APPLICANT: Carroll III, Eddie  
APPLICANT: Catino, Theodore J.  
APPLICANT: Dertl, Adnan  
APPLICANT: Ford, Donna M.  
APPLICANT: Lewis, Marcia E.  
APPLICANT: Monahan, John E.  
APPLICANT: Schlegel, Robert

TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION  
FILE REFERENCE: CCD-257 (US)  
CURRENT FILING DATE: 1999-06-08  
EARLIER FILING DATE: 1998-06-10  
NUMBER OF SEQ ID NOS: 850  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 516  
LENGTH: 450  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: (1)..(450)  
OTHER INFORMATION: n = A,T,C or G

US-09-328-111-516

Query Match 6.8%; Score 159.2; DB 4; Length 450;  
Best Local Similarity 69.8%; Pred. No. 4,1e-31;  
Matches 215; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY	975	GGGGGGGAGAGCAGCCGCGCTGTAAACCTCCAAAGAGACGTCGCGGACTGCACACA	1034
DB	122	GGTGAAGAGAGGTGTGCGCCCATCAAGCCCGCAAGAGCTTGTACTGCACACA	181
QY	1035	GCACCCAGCAGCAGAGAGAGCAGCAGAGTCTCGAGCAGCTCAGTCTCAGCGGCAT	1094
DB	182	ACACACGAGAGCTCTTAAGAAAGAAAGCTTTACAGAAAGTTAAACATTGCAATGGCAT	241
QY	1095	CCTCAAGAGAGTGTGGCAAGAGCAGCGCGCTTACGCTTGTCAACACCTGT	1154

DB 242 TTGGAAGAGTACTCTCTAAGAACATGCTGCTATGCTTGGCCTTTCTATAAACAGT 301  
QY 1155 GAGGTGGAGGACATGGGCTTACAGACTACTGTGACATCATCAAGCCCATGACAT 1214  
DB 302 GATGCTTCTGCACTTGGCTGCTGACTACATCATCATTAAGCAACCCATGACCT 361  
QY 1215 GACACAAATCAATGTAAGTGAAGGCGCCCTGACTGATCCGATGCTGAGAGATTGTGC 1274  
DB 362 CACACATGTAACCGGAAGATGAGAACCGTGTATTCACGGGATGCACGAGATTGCTGC 421  
QY 1275 TGAGCTCC 1282  
DB 422 TGATGTAC 429  
RESULT 2  
US-09-397-787-243  
Sequence 243, Application US/09397787  
Patent No. 6468758  
GENERAL INFORMATION:  
APPLICANT: Benson, Darin R.  
APPLICANT: Lodes, Michael J.  
APPLICANT: Mitcham, Jennifer L.  
APPLICANT: King, Gordon E.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR OVARIAN  
TITLE OF INVENTION: CANCER THERAPY AND DIAGNOSIS  
FILE REFERENCE: 210121.466C2  
CURRENT APPLICATION NUMBER: US/09/397,787  
CURRENT FILING DATE: 1999-09-16  
NUMBER OF SEQ ID NOS: 334  
SOFTWARE: FastSeq for windows Version 3.0  
SEQ ID NO 243  
LENGTH: 429  
TYPE: DNA  
ORGANISM: Homo sapien  
US-09-397-787-243  
Query Match 4.5%; Score 104.8; DB 4; Length 429;  
Best Local Similarity 72.3%; Pred. No. 2.6e-17;  
Matches 136; Conservative 0; Mismatches 52; Indels 0; Gaps 0;  
QY 1810 TATGAGTCGAGGAGAGGACAGTCAAGCTATGTCTATGAGAGAGCGGACGCTC 1869  
DB 242 TATGATTCGAGAGGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 301  
QY 1870 AGCTTGACATCAACAGCTCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1929  
DB 302 AGCTTGATATCAATAGTACTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 361  
QY 1930 TCACGGAGAGCCCTCCTGGAAGATTCCAAAGAGAGAGAGAGAGAGAGAGAGAG 1989  
DB 362 GCGAGGAGAGCCCTCTAGCTGATCAATCCAGAGAGAGAGAGAGAGAGAGAGAG 421  
QY 1990 CTGAGCC 1997  
DB 422 CTCAGCC 429  
RESULT 3  
US-08-232-463-14/c  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA

COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: PTZpPt-F15  
US-08-232-463-14  
Query Match 3.6%; Score 84.6; DB 1; Length 7218;  
Best Local Similarity 3.1%; Pred. No. 1.7e-11;  
Matches 12; Conservative 247; Mismatches 126; Indels 0; Gaps 0;  
QY 1508 GTTCAGCTGATGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1567  
DB 1438 GTACRR 1379  
QY 1568 TCAAGCCGTCGAG 1627  
DB 1378 RRR 1319  
QY 1628 AAGAGGAG 1687  
DB 1318 RRR 1259  
QY 1688 AAGAGATATAAAG 1747  
DB 1258 RRR 1199  
QY 1748 GCAGCAACCAATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1807  
DB 1198 RRR 1139  
QY 1808 CGTATGATGAG 1867  
DB 1138 RRR 1079  
QY 1868 TCAGCTTGACATCAACAGCTCC 1892  
DB 1078 RRRRRRRRRRRRRATGCAAGCTCC 1054  
RESULT 4  
US-08-257-073-4  
Sequence 4, Application US/08257073  
Patent No. 576597  
GENERAL INFORMATION:  
APPLICANT: Paoletti, Enzo











OY	654	GAATCTCTCTCTGTGAGGCACAGGCTACACCCCTTCCCTGGCGTACACCCGGAGCTCAT	713
Db	9948	AACCCAGCCCCCATACCAACCCACTGGCGACCTTGACCCGTGAACAACACCCCAACCCCA	9889
OY	714	CGTCCAGACCCCTGTCTCATGTAGAGTGGTGGCTTCCCGACGCATGTACAGAGCCCCCGCAGT	773
Db	9888	CGACCCCGGCACCCGATCTCCCCCGGACACACAGCACCGGGCCCCCGAAGACACACCCACAGA	9829
OY	774	GCCCCCCAGCACAAACCCCAACCCGCTCCAGCTTCCCAAGCCCGTACAGAGCACCCACC	833
Db	9828	GCCCCGAGCAGACACACATCATCAAGACCCGGCAGAACTCTCAAGGCCCGCGGCCACACAC	9765
OY	834	CATATATGGCGGCCACCCCAACAGCTTGTGAAGACAAGAAAGGAGTGAAGAGGAAGGAACAGA	893
Db	9768	ACCCGACGCTTACCGAAGCAGACGACGACCGACACACCAAGACACCCCGCCATCTCCGCGGA	9709
OY	894	CACGACACACCCCAACCACTTGAACCCCATTTACAGAGCAACCCCTGGCTGCCCGCGAACC	953
Db	9708	TCCGCAACCCACCGGCTCCACACACTACGACCAACCGACCGCTGGCGCCGCAACGCCCC	9649
OY	954	CAGACACCAACAG	966
Db	9648	TCGACCGCGCAG	9636

RESULT 15  
15-09-2004

```

1      Sequence 1, Application US/08804198
2      Patent No. 5945320
3      GENERAL INFORMATION:
4      APPLICANT: Burgett, Stanley G.
5      APPLICANT: Kuhnstoss, Stuart A.
6      APPLICANT: Rao, Nagaraaj R.
7      APPLICANT: Richardson, Mark A.
8      APPLICANT: Rosteck, Paul R., Jr.
9      TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENES
10     NUMBER OF SEQUENCES: 6
11     CORRESPONDENCE ADDRESS:
12     ADDRESSEE: PAUL R. CAMTRELL 1138
13     STREET: LILLY CORPORATE CENTER
14     CITY: INDIANAPOLIS
15     STATE: IN
16     COUNTRY: USA
17     ZIP: 46285
18     COMPUTER READABLE FORM:
19     MEDIUM TYPE: Floppy disk
20     COMPUTER: Macintosh
21     OPERATING SYSTEM: Macintosh 7.0
22     SOFTWARE: Microsoft Word 5.1
23     CURRENT APPLICATION DATA:
24     APPLICATION NUMBER: US/08/804,198
25     FILING DATE:
26     CLASSIFICATION: 435
27     ATTORNEY/AGENT INFORMATION:
28     NAME: CAMTRELL, PAUL R.
29     REGISTRATION NUMBER: 36,470
30     REFERENCE/DOCKET NUMBER: P9113
31     TELECOMMUNICATION INFORMATION:
32     TELEPHONE: 317-276-3885
33     INFORMATION FOR SEQ ID NO: 1:
34     SEQUENCE CHARACTERISTICS:
35     LENGTH: 44377 base pairs
36     TYPE: nucleic acid
37     STRANDEDNESS: single
38     TOPOLOGY: linear
39     MOLECULE TYPE: DNA (genomic)
40     FEATURE:
41     NAME/KEY: CDS
42     LOCATION: 350..14002
43     FEATURE:
44     NAME/KEY: CDS
45     LOCATION: 14046..20036
46     FEATURE:

```

```

? NAME/KEY: CDS
? LOCATION: 20110..31284
?
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 31329..36071
?
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 36155..41830
?
US-08-804-198-1

```

Query Match	Score	DB 2;	Length
2.68;	59.4;	DB 2;	44377;

Matches 177; Conservative 0; Mismatches 196; Indels 0; Gaps 0;

OY	594	TTCCAGGTTCAAAACACACATCTAAGATGATTCCTCCGACAGCCACAGACCCCTCACCC	653
Db	10008	TTCCCAACACACCGACACACACTTATCTACGACTCTCGGAAAACCCCACTCCCGCATAC	9949
OY	654	GAATCTCTCTCTGTGAGCGCACGGCTACCCCTTCCCTGCCGTACCCCGGACCTCAT	713
.Db	9948	AACCCAGCCCCATACCAACCCACTGGCGACCTGTACCCGTGAACAACCCCCACCCCA	9889
.OY.	714	CGTCAGACCCCTCTATGACAGTGGTGCTCCCAAGCCACTGTGACAGCGCCCGGCAGT	773
Db	9888	CGACCCCGCACCCGCACTCCCCCGCACACAGCACCGGCCCGGAAAGACACACCCACAGA	98239
OY	774	GGCCCCAGGCACAAACCCCCACCGCGCTCAGTCTCCCAAGCGCCGTACAGAGCAACCAAC	833
Db	9828	GGCCCCGAGCCACACACATCTCAAGACCGCGCAACTCTTCAAGCGCCCGGCCACCAAC	9769
OY	834	CATCATCGCGCGCCACCCCAACGCTGTGAAGACAAAGAGGAGTGAAGAGGAAGAAGCA	893
Db	9768	ACCGAGAGCTACCGGAACGACGACGAGAGACCAACCAAGACCCCGCCACATCCCGCGA	9709
OY	884	CACCAACACCCCCACACCACTTGAACCCATTACAGAGCCACCTTGCTGCCCCGGAGCC	953
Db	9708	TTCGCAACCAACCCCTTCCACACACTCAGCAAGCAACCGACCGCTGCGCCGCAACGCCCC	9649
OY	954	CAAGACCAACCAAG	966
Db	9648	TTCGACCGCGAG	9636

```
Search completed: March 1, 2003, 14:44:56
Job time : 244 secs
```

Job time : 244 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 1, 2003, 11:57:34 ; Search time 140 Seconds  
(without alignments)  
10374.344 Million cell updates/sec

Title: US-09-700-590A-101  
Perfect score: 2329  
Sequence: 1 gcaggagatcactagcatgctc.....ngccttaaaagttnccca 2329

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapept 1.0

Searched: 460893 seqs, 311809382 residues

Total number of hits satisfying chosen parameters: 921786

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_NA.\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	460.2	19.8	4664	10	US-09-880-107-1657 Sequence 1657, App
2	224	9.6	424	10	US-09-764-864-785 Sequence 785, App
3	224	9.6	746	10	US-09-764-864-372 Sequence 372, App
4	159.2	6.8	450	10	US-09-879-536-516 Sequence 516, App
5	137	5.9	358	10	US-09-783-590-1966 Sequence 1966, App
6	104.8	4.5	429	10	US-09-876-889-243 Sequence 243, App
7	66.6	2.9	244	10	US-09-925-301-299 Sequence 299, App
8	64.8	2.8	8147	12	US-10-109-886-9 Sequence 9, App
9	64	2.7	463	10	US-09-960-352-7186 Sequence 7186, App
10	63.6	2.7	681	10	US-09-822-830A-26 Sequence 26, App
11	63.6	2.7	12733	9	US-10-032-393-47 Sequence 47, App
12	63.6	2.7	12739	9	US-10-032-393-8 Sequence 8, App
13	63.4	2.7	8895	10	US-09-764-853-887 Sequence 887, App
14	63.4	2.7	8895	10	US-09-764-853-937 Sequence 937, App
15	63.4	2.7	9656	10	US-09-764-853-886 Sequence 886, App
16	63.4	2.7	9656	10	US-09-764-853-933 Sequence 933, App
17	62	2.7	423	10	US-09-864-761-18355 Sequence 18355, A
18	61.4	2.6	283	10	US-09-960-352-9095 Sequence 9095, App
19	61	2.6	217	10	US-09-960-352-6539 Sequence 6539, App

20	61	2.6	2269	9	US-09-989-920-113 Sequence 113, App
C 21	60.8	2.6	341	10	US-09-960-352-12302 Sequence 12302, A
C 22	60.6	2.6	446	10	US-09-960-352-3400 Sequence 3400, App
C 23	60.2	2.6	415	10	US-09-960-352-2223 Sequence 2223, App
C 24	60.2	2.6	2608	10	US-09-739-254-16 Sequence 16, App
C 25	60.2	2.6	2608	10	US-09-904-615-16 Sequence 16, App
C 26	59.4	2.6	378	10	US-09-960-352-14879 Sequence 14879, A
C 27	59.4	2.6	475	10	US-09-864-761-6203 Sequence 6203, App
C 28	59.4	2.6	511	10	US-09-864-761-22817 Sequence 22817, A
C 29	59.4	2.6	2903	9	US-09-822-846-739 Sequence 239, App
C 30	59.2	2.5	547	10	US-09-815-343-213 Sequence 213, App
C 31	58.6	2.5	305	10	US-09-864-761-19262 Sequence 19262, A
C 32	58.6	2.5	496	10	US-09-864-761-2534 Sequence 2534, App
C 33	58.4	2.5	261	10	US-09-864-761-19524 Sequence 19524, A
C 34	58.4	2.5	261	10	US-09-864-761-25517 Sequence 25517, A
C 35	58.4	2.5	448	10	US-09-864-761-557 Sequence 557, App
C 36	58.4	2.5	451	10	US-09-860-352-10262 Sequence 10262, A
C 37	58.4	2.5	513	10	US-09-864-761-8851 Sequence 8851, App
C 38	58.4	2.5	4237	10	US-09-745-763-20 Sequence 20, App
C 39	58.2	2.5	276	10	US-09-864-761-20595 Sequence 20595, A
C 40	58.2	2.5	529	10	US-09-983-965-2109 Sequence 2109, App
C 41	58.2	2.5	12980	9	US-09-985-937-5 Sequence 5, App
C 42	58.2	2.5	12980	10	US-09-238-076-5 Sequence 5, App
C 43	57.6	2.5	385	10	US-09-960-352-6420 Sequence 6420, App
C 44	57.6	2.5	425	10	US-09-960-352-4010 Sequence 4010, App
C 45	57.6	2.5	456	10	US-09-960-352-9659 Sequence 9659, App

## ALIGNMENTS

RESULT 1  
US-09-880-107-1657  
Sequence 1657, Application US/09880107  
Patent No. US2002042981A1  
GENERAL INFORMATION:  
APPLICANT: Horne, Darci T.  
APPLICANT: Vockley, Joseph G.  
APPLICANT: Scherf, Uwe  
APPLICANT: Gene Logic, Inc.  
TITLE OR INVENTION: Gene Expression Profiles in Liver Cancer  
FILE REFERENCE: 44921-5028-NO  
CURRENT APPLICATION NUMBER: US/09/880,107  
CURRENT FILING DATE: 2001-06-14  
PRIOR APPLICATION NUMBER: US 60/211,379  
PRIOR FILING DATE: 2000-06-14  
PRIOR APPLICATION NUMBER: US 60/237,054  
PRIOR FILING DATE: 2000-10-02  
NUMBER OF SEQ ID NOS: 3950  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1657  
LENGTH: 4664  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: Genbank Accession No. US2002042981A1 D42040  
US-09-880-107-1657  
Query Match 19.8%; Score 460.2; DB 10; Length 4664;  
Best Local Similarity 57.0%; Pred. No. 1.1e-101;  
Matches 1151; Conservative 0; Mismatches 693; Indels 175; Gaps 10;  
143 CCACCCCCCCCCCGAGAGCCGCAACCTTACCAAGCCCAAGAGCGAGCAACCAAC 202  
1877 CCAACCCACACCCCCCGAGAGTGTTCATTCCAAAAGCCGAGAGCTTACCAACGCG 1936  
203 TGCATATCCTGCTGAGATGCTGCTCAAGACATATGAAACACACAGTTGCGAGCCTT 262  
1937 TGCATATCCTGCAACAGTATGATGAAGGCTCTGTGGAACATCAGTTCGATGCGCAT 1996  
263 TCCAGCAGCCTGTGATGCGCTCAAGCTGAACCTCCCTGATTTATTAAGTCTTAA 322  
1997 TCCGAGAGCCTGTGATGCTGTCAAACTGGCTTACCGGATTAACCAAAATTTAAAC 2056

```
QY 323 CCGCTATGATATGGACAAATTAAGAGCGCTTGGAAAAACAATATTAAGTATGCTC 382
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2057 ACCCTATGATATGGCTACTATTAAGAGGAGACTTGAACAAATATTAATGGCTGCTT 2116
QY 383 AGGAATGATCCAGAGCTCAACACTATGTTTCAAAATGTTTCACTACAAACCTG 442
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2117 CAGAGGATATGCAAGATTTTAAATACATGTTCCAACTGTTACATTTTCAACAAACCCA 2176
QY 443 GAGATGACATTAATGATGAGAGCTTGGAAAAAGCTCTTCTGCAAAAAATTAATG 502
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2177 CGATGATATTAATGATGAGCAAAACGCTGAAAAAGATTTCTACAGAAAGTTGCAT 2236
QY 503 AGCTACCCACAGAAACCGAGATCATGATCTCAGGCAAAAAGAGAGAGAGCTGGGA 562
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2237 CATGCGACACAGAAAGAAAGAGCTGTAGTACCATCCCTTAAGAACAGCCACAAAGAG 2296
QY 563 GGAAGAAACAGGAGACAGCAAAACCTGGGGTTTCCAGGGTACCAACACACAGCAT 622
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2297 GGGCCAAAGTTGGCAGCGCTCCAGGGCAGTGTACAGTGCCTCAGTGCCTGCGCTCT 2356
QY 623 CGACTCTCCGAGACACCCCTCAGCCGATCTCTCTCTCTGTCAGGCGCAGCTC 682
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2357 CTTCTGTGTACACACAGCCCTGTATATCTCTCCAGCTGAGATACCTACACACTG 2410
QY 683 ACCCTTCCCTGCGTCCAGCCCGGAGCTCATCTCCAGACCCCTGTCAATGACAGTGGTG 742
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2411 TCCTCAACATTCCTCCACCCCATCATGATCTTTC-----T 2443
QY 743 CTCGCCAGCAGCTGACAGAGCCCGGCAAGTCCCGCCCAAGCCCAACACCCCGGCTC 802
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2444 CTCACATTTCTCAAGTCTTGTGACCTGTGTGAGACCCCGCTCTGCTGTATGTCAGCTC 2503
QY 803 CAGCTTCCAGCCCGCTGACAGAGCCACCCATCATCTGCGGCGCCACCCAGCTGTGGA 862
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2504 CTCACAGCCAGCCCGCTTGGCAAGAAAAAGGCGT----- 2537
QY 863 AAGCAAAAGAGAGTGAAGAGAAAGAGACACACACACCCACCACCATTTGACCCCA 922
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2538 ----AAAGGGAAAGAGATTAACACCCCTACACCTACAGCCTTGGCTCTGCTG 2593
QY 923 TTCACAGAGCAGCCCTGCTGCCCCCGAGGCCCAAGACCAACAGCTGGGGC---AGCGG 979
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2594 CTCACAGCTAGCCCTCTCTGGGAGCTTGTAGGCTTAAGGACAGAGGCTTCCCTATGCGTA 2653
QY 980 GGGAGAGACAGCGGCTGTGAAACCTCAAAAGAGAGCTGCCGACTCTCAGACAGCAC 1039
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2654 GAGAGAGTGTGGCCCTCAAGCCCAAGCAAGAGCTTGGCTGACTCTCAGCAACAC 2713
QY 1040 CAGCAACCAAGAAAGACACAGGTCTCGGAGAGCTCAAGTGTCTCAGCGGATCTCTCA 1099
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2714 ACACAGCTCTAAGAAAGAAAGCTTTCAGAAAGTTAAACATTTGCAATTTGGA 2773
QY 1100 AGGAGATGTTTGCAGAAAGACAGCGCCCTACGCTGGCCCTTCAAGCCGTGGAGC 1159
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2774 AGGAGTACTCTCTAAGAAAGCATGCTGCTATGCTTGGCTTCTTAAACAGTGGAGT 2833
QY 1160 TGGAGGACTGGGCTTACACAGCTACTGTGACATCATCAAGACCCCATGAGCATGACA 1219
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2834 CTTTGTCACTTGGCCGTGACATCATCAATTAAGACCCCATGAGACTCAGCA 2893
QY 1220 CAATCAAGCTTAACTGAGAGCCCGGAGTACCGTATGCTCAGGAGTTTGGTGTGAG 1279
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2894 CGTGTCAAGGGAGAGATGAGAAACCGTATTAACCGGATTCACAGGAGTTTGGCTGAGT 2953
QY 1280 TCCGATGATGTTTCTCACTGCTATTAAGTACAAACCTCTGACATGAGTGGTGGCA 1339
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2954 TACGGCTTATGTTCTCACTGCTATTAAGTACATCCCAAGTCCAGTGTGTGGCAA 3013
QY 1340 TGGCCCGCAGCTCCAGATGTTTGAATGCGCTTTCAGAGATGCGGAGAGAGCTG 1399
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3014 TGGCAGAAAGAGTACAGATGATATTGAGTTCGGTATGCAAGATGCCAGATGAACAC 3073
```

```
QY 1400 AGGAGCAGTGTGTGCGCTGTCTCCCGGCAAGTCCCTCC-----ACCAAGTTGTG 1456
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3074 TAGAACCAAGGGCTTTACAGATCTTACTGACATGCCCCCTGGCTGGCCAAATCGCTT 3133
QY 1457 CCCCCGCCATCCAGCAGCAGCAGCAGGATAGCTCTC----- 1496
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3134 CAGAGTCTTCAGTGAAGAAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3193
QY 1497 -----GCAAGTGAAGTTCAGTGTGAGTGTGAGTGTGAGAGAGAGAGAGAGAGAG 1540
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3194 ATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3253
QY 1541 AGCGCTGTGCTGATCTCCAGAGACAGCTCAAAACCGTGAAGAGAGAGAGAGAGAGAG 1600
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3254 ATCGCTTACAGAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3313
QY 1601 CTCAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1659
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3314 CCGAGGGTCCATATATCAAGAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3373
QY 1660 GAAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1705
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3374 GGAAGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3433
QY 1706 AAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1749
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3434 CACCCCGCCAGCTCAACCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3493
QY 1750 -----AGCAAGCAGATGTGAGCAAGA 1771
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3494 CTGCTTATAGCCCTTGTGCTTGTGAGCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 3553
QY 1772 AGGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1831
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3554 AGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3613
QY 1832 AGTGAAGCTATGCTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1891
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3614 AGAGCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3673
QY 1892 CCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1951
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3674 CTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3733
QY 1952 ATTCAAGCCCGAGAGATTTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2011
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3734 ATTCAAGCCCGAGAGAGAGAGATTTGAGATGATTTGAAGACACTCAAGCCATCAAGAG 3793
QY 2012 AGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2050
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3794 AGCTT- GAGCGCTATGCTTCTTCTGCTTCAAGAGAA 3831
```

## RESULT 2

US-09-764-864-785

Sequence 785, Application US/09764864

Patent No. US20020132753A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic acids, proteins, and antibodies

FILE REFERENCE: PT223

CURRENT APPLICATION NUMBER: US/09/764,864

CURRENT FILING DATE: 2001-01-17

Prior application data removed - consult PALM or file wrapper

NUMBER OF SEQ ID NOS: 1792

SOFTWARE: Patentln Ver. 2.0

SEQ ID NO 785

LENGTH: 424

TYPE: DNA

ORGANISM: Homo sapiens

US-09-764-864-785

Query Match 9.6%; Score 224; DB 10; Length 424;



Best Local Similarity 74.0%; Pred. No. 6.9e-45;  
Matches 284; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

QY	165	CTCCACCCCTAAACAAGCCCAAGSAGSAGCAACCAACCACTGCAATACCTGCACAGATGGT	22
Db	2	CTCCACCCCAAGCAAGCCCGCCGCCAAGACCAACCACTGGAGTACATGCAGAAATGGGT	61
QY	225	GCTCAAGACACTATGGAAMACACAGTTTGCATGGCTCTTCAGCAGCGCTGTGATGGCGT	28
Db	62	GGTGAAGACGCTGTGGAACACAGCTGGCTGGCCCTTACACAGCCGCGAGACGAAT	122
QY	285	CAAGCTGAACCTCCCTGATTATCTAAAGATCATTTAAACGCTAAVGGATATGGACAAT	344
Db	122	CAATTGCAACCTGCCGATTTATCTAAATTAATTAACCAACCATGGATATGGGACTAT	181
QY	345	AAAGAAAGCGCTTGGAAAAACAACCTTTACTGGAATGCTCAGGAATGTATCCAGACTTCAA	40
Db	182	TAAAGAAAGACTAAGAAAAATTTTATTTGGAGTGCAGAGGAAGTATGAGAGCACTTCAA	241
QY	405	CACATGCTTTCAAAATTTGTTCATCTCAACAACCCCTGGAGATGACATAGCTTAATGGC	464
Db	242	CACCATGTTTCAAAATTTGTTCATTTTAAACAAAGCCACAGATGACATAGTGTCTAATGGC	301
QY	465	AGAAAGCTCTGGAAAAAGCTCTTTCGAAAAAATTAATGAGCTACCAACAGAAACCGA	524
Db	302	CCAAGCTTTAGAGAAATTTTTCACAAAAAAGTGCCCAAGATGCCCAAGAGGAAGTTGA	361
QY	525	GATCATGTATAGTCCAGGCAAAAAGG	548
Db	362	AATTTATACCCCTGCCTCCAAAGG	385

RESULT 3  
US-09-764-864-372

```

Sequence 372, Application US/09764864
Patent No. US20020132753A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: P1723
CURRENT APPLICATION NUMBER: US/09/764,864
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 1792
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 372
LENGTH: 746
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (387)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (646)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (665)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (736)
OTHER INFORMATION: n equals a,t,g, or c
US-09-764-864-372

```

Query Match	9.68; Score 224; DB 10; Length 746;
Post Total	74.00

	Matches	284;	Conservative	0;	Mismatches	100;	Indels	0;	Gaps
0y	165	CTTCAAGCCCTAAACAGGCCAAGAGCCCAACCACTCATTACCTGCTCAGAGTGT	224						
Db	2	CTCAACACCCGACGAGCCCGGCCGCAAGACCAACCACTGATGATACATGCAAGATGTGT	61						
0y	225	GCTCAAGACACTATGGAACAACCAAGCTTTTCAGAGGCCCTTGAGATGGCGT	284						

[illegible]

RESULT 4  
US-09-879-536-516

```

? Sequence 516, Application US/09879536
? Patent No. US20020144298A1
? GENERAL INFORMATION:
? APPLICANT: Endeavor, Wilson O.
? APPLICANT: Steilmann, Kathleen E.
? APPLICANT: Astle, Jon H.
? APPLICANT: Burgess, Christopher C.
? APPLICANT: Bushnell, Steven E.
? APPLICANT: Carroll III, Eddie
? APPLICANT: Catino, Theodore J.
? APPLICANT: Dettl, Adnan
? APPLICANT: Ford, Donna M.
? APPLICANT: Lewis, Marcia E.
? APPLICANT: Monahan, John E.
? APPLICANT: Schlegel, Robert
? TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
? TITLE OF INVENTION: PRODUCTS
? FILE REFERENCE: CCD-257 (US)
? CURRENT APPLICATION NUMBER: US/09/879,536
? CURRENT FILING DATE: 2001-09-21
? PRIOR APPLICATION NUMBER: US 60/088,801
? PRIOR FILING DATE: 1998-06-10
? NUMBER OF SEQ. ID NOS: 850
? SOFTWARE: FastSeq for Windows Version 3.0
? SEQ ID NO 516
? LENGTH: 450
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (1)...(450)
? OTHER INFORMATION: n = A,T,C or G
? US-09-879-536-516

```

Query Match	6.8%	Score 159.2	DB 10	Length 450
Best Local Similarity	69.8%	Pred. NO. 3.8e-29		
Matches 215; Conservative	0	Mismatches 93	Indels 0	Gaps 0

QY	975	GGCGGGGAGGACGCGCGGCTGTGAAACCTCCAAAGGAGAGCGGCCGATCTCAGCA	103
Db	122	GGGTAGAGAGAGTGTGTGCGCCCATCAAGCCGCCCAAGACTTCCGACTCTCAGCA	181
QY	1035	GCACCCGACGACACAGAGAAAGGACGACAAAGTCTGTGGAGCAGCTCAGTGTGCTGAGCGGCAT	109
Db	182	ACAACACACGAGAGCTCTTAAGAAAGAAAGCGCTTGTGAAACGTTTAAACATTTGAAAGAGGAT	241
QY	1095	CCCTCAAGAGATGTTTCCCAAGAGACGCGCGCTACGCGCTGGCCCTTCTACAACTGT	115

```
Db 242 TTTGAAGGACTTACTCTCTAAGAGCATGCTGCTATGCTTGGCTTTCTATTAACCACT 301
OY 1155 GGACGTGGAGGACACGTGGCCCTACACGACTGATCATCATCAAGACCCCATGACAT 1214
Db 302 GGATGCTTCTGACCTGCTGCTGATCATCATCATTAAGACCCCATGACACT 361
OY 1215 GAGCAATCAATCTTAATCTGAGAGCCCTGATGATCCGATGATGATGATGATGATGAT 1274
Db 362 CAGCACTGCTCAACGGAGATGAGAACCTGATTAACCGGATGACAGAGATTTGCTGC 421
OY 1275 TGAGCTCC 1282
Db 422 TGATGTAC 429

RESULT 5
US-09-783-590-1966
; Sequence 1966, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16,2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1966
; LENGTH: 358
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (10)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (41)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (102)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (121)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (133)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (158)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (180)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (186)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (203)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (225)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (237)
```

```
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (329)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (331)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (334)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (336)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (339)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (347)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (355)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-1966
```

Query Match 5.9%; Score 137; DB 10; Length 358;  
Best Local Similarity 82.3%; Pred. No. 8,3e-24;  
Matches 237; Conservative 0; Mismatches 39; Indels 12; Gaps 7;

```
OY 881 AGAGAAAGACGACACACACACACACACACACACACACACACACACACACACACACACAC 940
Db 6 AGGNNAAAGACGACACACACACACACACACACACACACACACACACACACACACACACAC 65
OY 941 TGCCCCGGAGCCCAAGACACCAAGCTGGCCAGCGGGAGAGACACCGGCTGTG- 999
Db 66 TGCCCCGGAGCCCAAGACACCAAGCTGGCCAGCGGGAGAGACACCGGCTGTG- 125
OY 1000 --AAACCTCAAGAAGAGAGCTGCCGACTCT-CAGCAGCACCACACACACAGAA--GA 1054
Db 126 GAAACCTNCCAAAGAGAGAGCTGCCGACTCTCAGCAGCACCACACACAGAGAAAGAG 185
OY 1055 GCAGCAAGGCTCGAGAGC--AGCTCAAGTGTCTCAGCGG-CATCTCAAGGA-GATGTTT 1110
Db 186 NCAGCAAGGCTCGAGAGCAGCTTCAAGTGTCTCAGCGGNCATCTCAAGGAGATGTTT 245
OY 1111 GCCAAGACGACGCGCTTACGCT--GGCCCTTCTCAAGCCTGTGG 1156
Db 246 GCCAAGACGACGCTTACGCTTGGCCCTTCTCAAGCCTGTGG 293
```

```
RESULT 6
US-09-876-889-243
; Sequence 243, Application US/09876889
; Patent No. US2002007615A1
; GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR OVARIAN
; FILE REFERENCE: 210121.466C3
; CURRENT APPLICATION NUMBER: US/09/876,889
; CURRENT FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 353
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 243
; LENGTH: 429
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-876-889-243
```

Query Match 4.5%; Score 104.8; DB 10; Length 429;  
Best Local Similarity 72.3%; Pred. No. 6e-16;





```
OY 622 TCAGCTCTCCGCAAGACCAGACCCCTCAGCCGAATCTCTCTGTGACAGCCAGCCT 681
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5358 TCGACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 5299
OY 682 CACCCCTTCCCGCGCTACACCCCGGACTCATGTGTCAGACCCCTGTATGACAGTGTG 741
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5298 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 5239
OY 742 CCTCCCGACGCTGACAGACGCCCGCCGCAAGTGGCCCCCGACCAACCCCGACCCCT 801
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5238 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 5179
OY 802 CCACTCCCGACGCGGACAGACGACCCACCATCATGTGGGAGCCACCCAGAGCTGTG 861
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5178 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 5119
OY 862 AAGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 921
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5118 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 5059
OY 922 ATTACGAGACGACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 963
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5058 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 5017
```

```
RESULT 13
US-09-764-853-887/c
; Sequence 887, Application US/09764853
; Patent No. US20020090672A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ206
; CURRENT APPLICATION NUMBER: US/09/764,853
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 887
; LENGTH: 8895
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-853-887
```

```
Query Match 2.7%; Score 63.4; DB 10; Length 8895;
Best Local Similarity 52.9%; Pred. No. 3.3e-05;
Matches 136; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

OY 1610 AGCAGAACAAACCAAGAAAAAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1669
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1004 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 945
OY 1670 AAAGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1729
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 944 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 885
OY 1730 AGACGAAGAAAAATATATAGCAGCAATGTGAGCAAGAGAGAGAGAGAGAGAGAG 1789
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 884 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 825
OY 1790 AGACGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1849
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 824 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 765
OY 1850 ATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1866
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 764 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
```

```
RESULT 14
US-09-764-853-937/c
; Sequence 937, Application US/09764853
```

```
; Patent No. US20020090672A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ206
; CURRENT APPLICATION NUMBER: US/09/764,853
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 937
; LENGTH: 8895
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-853-937
```

```
Query Match 2.7%; Score 63.4; DB 10; Length 8895;
Best Local Similarity 52.9%; Pred. No. 3.3e-05;
Matches 136; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

OY 1610 AGCAGAACAAACCAAGAAAAAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1669
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1004 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 945
OY 1670 AAAGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1729
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 944 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 885
OY 1730 AGACGAAGAAAAATATATAGCAGCAATGTGAGCAAGAGAGAGAGAGAGAGAGAG 1789
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 884 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 825
OY 1790 AGACGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1849
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 824 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 765
OY 1850 ATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1866
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 764 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
```

```
RESULT 15
US-09-764-853-886/c
; Sequence 886, Application US/09764853
; Patent No. US20020090672A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ206
; CURRENT APPLICATION NUMBER: US/09/764,853
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 886
; LENGTH: 9656
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-853-886
```

```
Query Match 2.7%; Score 63.4; DB 10; Length 9656;
Best Local Similarity 52.9%; Pred. No. 3.4e-05;
Matches 136; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

OY 1610 AGCAGAACAAACCAAGAAAAAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1669
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1761 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1702
OY 1670 AAAGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1729
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1701 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1642
OY 1730 AGACGAAGAAAAATATATAGCAGCAATGTGAGCAAGAGAGAGAGAGAGAGAGAG 1789
```



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

```
Run on:      March 1, 2003, 11:52:00 ; Search time 3254 Seconds
              (without alignments)
              11591.668 Million cell updates/sec
```

Title:	US-09-700-590A-101
Perfect score:	2329
Sequence:	1 gcagagatcactagcatgtc.....ngccttaagaagttnccca 2329

Scoring table: IDENTITY\_NDC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 32308132

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

```
Database :

EST:*
1: em_estba:*
2: em_esthun:*
3: em_estln:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_qrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other
26: em_gss_pro:*
27: em_gss_rtd:*
```

**Pred. NO.** is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result	No.	Score	Query Match	Length	DB	ID	Description
1	1656.8	71.1	1716	11	BC000156	BC000156	Homo sapi
2	1640	70.4	1743	11	BC030158	BC030158	Homo sapi
3	1411	60.6	1776	11	BC011541	BC011541	Mus muscul
4	856	36.8	987	14	B0651563	B0651563	ACENOCOURT
5	787	33.8	951	14	B0645784	B0645784	ACENOCOURT
6	779	33.4	893	14	B0219777	B0219777	ACENOCOURT

7	770.2	33.1	93.3	14	B0214403	B0214403	AGENCOURT
8	756.6	32.5	97.5	14	B065387	B065387	AGENCOURT
9	752.4	32.3	92.5	14	B0645818	B0645818	AGENCOURT
10	736.2	31.6	89.9	14	B0647311	B0647311	AGENCOURT
11	735.8	31.6	93.1	14	B0650713	B0650713	AGENCOURT
12	733.4	31.5	91.9	14	B0652326	B0653326	AGENCOURT
13	730.4	31.4	92.6	14	B0650970	B0650970	AGENCOURT
14	729	31.3	74.0	14	BM983041	BM983041	UT-CF-ENI
15	715.2	30.7	93.8	14	B0654353	B0654353	AGENCOURT
16	687.8	29.5	110.6	13	BM452148	BM452148	AGENCOURT
17	665.4	28.5	208.5	11	BC010699	BC010699	Homo sapi.
18	664.2	28.5	97.1	14	B0652321	B0653321	AGENCOURT
19	647.8	27.1	81.5	14	B0749323	B0749323	UT-M-FCO-
20	630	27.1	105.5	12	B0624322	B0624322	602727345
21	619	26.6	64.1	13	B1907231	B1907231	603065284
22	613.8	26.4	62.7	14	BM679991	BM679991	UT-E-EO1-
23	611.6	26.3	90.2	14	B0232371	B0233371	AGENCOURT
24	606.8	26.1	66.6	12	BF118610	BF118610	K8781 no
25	605	26.0	70.9	13	B1414784	B1414784	6029911866
26	598.6	25.7	101.1	11	BC013666	BC013666	Mus muscu
27	598.4	25.7	61.9	14	B0939643	B0939643	AGENCOURT
28	594.8	25.5	66.2	13	B1252631	B1252631	602952637
29	593	25.4	61.5	12	BG391946	BG391946	602409883
30	592	24.6	58.3	14	B0674079	B0674079	AGENCOURT
31	565.6	24.3	57.2	9	A1394011	A1394011	tg11e04.x
32	552	23.7	64.5	10	BE294756	BE294756	601173993
33	549.4	23.6	93.7	12	BG032535	BG032535	602301448
34	547.6	23.5	67.5	12	BF550759	BF550759	UT-R-CO-j
35	547	23.5	95.0	12	BF530953	BF530953	602072464
36	544.8	23.4	56.8	9	A1480003	A1480003	tm171607.x
37	542.6	23.3	54.8	13	BM389037	BM389037	UT-R-D20
38	536.4	23.0	63.1	12	BF712114	BF712114	MT-P-H3-a
39	533.2	22.9	78.6	13	B1852791	B1852791	603379265
40	529.2	22.7	91.9	12	BF180998	BF180998	601805044
41	527.8	22.7	94.4	9	AL044223	AL044223	DRF2Pd34K
42	527.8	22.7	54.4	9	B0049232	B0049232	AGENCOURT
43	527	22.6	53.5	9	A1360247	A1360247	qy44f05.x
44	521.4	22.4	64.7	12	BF237405	BF237405	602025085
45	520.6	22.4	164.1	11	BC004832	BC004832	Mus muscu

## ALIGNMENTS

RESULT 1	BC000156		
LOCUS	BC000156	1716 bp	mRNA
DEFINITION	Homo sapiens, Similar to bromodomain-containing 4, clone		HTC 12-JUL-2000
ACCESSION	IMAGE:2900081, mRNA.		
VERSION	BC000156		
KEYWORDS	BC000156.1	GI:14704448	
SOURCE	HTC.		
ORGANISM	Homo sapiens.		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 1716)		
TITLE	Strausberg, R.		
JOURNAL	Direct Submission		
	Submitted (03-NOV-2000) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		

Contact: villalob@bcm.tmc.edu.  
Villalob, D.K., Luna, R.A., Hale, S.M., Huiy, S., Lu, X., Garcia,  
A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,  
Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found  
through the T.M.A.G.E. Consortium/BLN at: <http://image.llnl.gov>  
Series: IRK Plate: 3 Row: e Column: 7  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 3115203  
This clone has the following problem: frame shifted.

## FEATURES

Location/Qualifiers  
1..1716  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2900081"  
/issue\_type="Placenta, choriocarcinoma"  
/clone\_id="NH\_MGC\_10"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
BASE COUNT 474 a 549 c 427 g 266 t  
ORIGIN

Query Match 71.1% Score 1656 8: DB 11: Length 1716:

Best Local Similarity 99.9% Pred. No. 1.9e-287;

Matches 1658: Conservative 0: Mismatches 2: Indels 0: Gaps 0:

```

QY 4 GGGACTACTAGCATGTCGTGGAGAGCGGCCCTGGGACGAGATTGGAATCTGCCAGTA 63.
DB 57 GGGATCCTAGCATGTCGTGGAGAGCGGCCCTGGGACGAGATTGGAATCTGCCAGTA 116
QY 64 ATGGGGATGGACTAGAAATCTCCAAATGTCTCAACACAGGCCAGGCCCAACCCAG 123
DB 117 ATGGGGATGGACTAGAAATCTCCAAATGTCTCAACACAGGCCAGGCCCAACCCAG 176
QY 124 CCAGCCACGACGACGACCAACCCCGCCCGCCAGAGACCTCCAAACCTTACAAAGCC 183
DB 177 CCAGCCACGACGACGACCAACCCCGCCCGCCAGAGACCTCCAAACCTTACAAAGCC 236
QY 184 AAGAGGACGACCAACCACTGCAATACCTGCTCAGAGTGTCTCAAGACTATGAGAA 243
DB 237 AAGAGGACGACCAACCACTGCAATACCTGCTCAGAGTGTCTCAAGACTATGAGAA 296
QY 244 CACGACTTGGCATGGCTTCCAGCAGCGTGTGATGCCGTCAAGCTACCTCCCTGAT 303
DB 297 CACGACTTGGCATGGCTTCCAGCAGCGTGTGATGCCGTCAAGCTACCTCCCTGAT 356
QY 304 TACATATAGATCATTTAAAGCGCTATGATGATGGAACAATAAGAGCGCTGGAAC 363
DB 357 TACATATAGATCATTTAAAGCGCTATGATGATGGAACAATAAGAGCGCTGGAAC 416
QY 364 AACATATAGATCATTTAAAGCGCTATGATGATGGAACAATAAGAGCGCTGGAAC 423
DB 417 AACATATAGATCATTTAAAGCGCTATGATGATGGAACAATAAGAGCGCTGGAAC 476
QY 424 TACATATAGATCATTTAAAGCGCTATGATGATGGAACAATAAGAGCGCTGGAAC 483
DB 477 TACATATAGATCATTTAAAGCGCTATGATGATGGAACAATAAGAGCGCTGGAAC 536
QY 484 TACATATAGATCATTTAAAGCGCTATGATGATGGAACAATAAGAGCGCTGGAAC 543
DB 537 TACATATAGATCATTTAAAGCGCTATGATGATGGAACAATAAGAGCGCTGGAAC 596
QY 544 AAGAGGACGACGATGGGAGGAAAGAGAGGAGCAAAACCTGCTTCCAGCGTA 603
DB 597 AAGAGGACGACGATGGGAGGAAAGAGAGGAGCAAAACCTGCTTCCAGCGTA 656
QY 604 CCAAGACACACTCAGCATGCTCTCCGAGAGACCCAGACCCCTGAGCGAATCTCTCT 663
DB 657 CCAAGACACACTCAGCATGCTCTCCGAGAGACCCAGACCCCTGAGCGAATCTCTCT 716
QY 664 CCGTGTGACAGGACGCTGACCCCTGCTGCGCGTCAACCCGAGACTCATCGTCAGACC 723

```

```

DB 717 CCGTGTGACAGGACGCTGACCCCTTCCTCCGTCACCCCGAGACTCATGTGCAGACC 776
QY 724 CCGTGTGACAGGACGCTGACCCCTTCCTCCGTCACCCCGAGACTCATGTGCAGACC 783
DB 777 CCGTGTGACAGGACGCTGACCCCTTCCTCCGTCACCCCGAGACTCATGTGCAGACC 836
QY 784 CCAAAACCCCGAGCTGACCTCCAGCCCGTACAGAGCCAGCCAGCCAGCTATGCGG 843
DB 837 CCAAAACCCCGAGCTGACCTCCAGCCCGTACAGAGCCAGCCAGCCAGCTATGCGG 896
QY 844 GCGACCCGAGCGCTGGAAGCAAGAGGAGGTAAGAGGAGCAACAGACACACACACC 903
DB 897 GCGACCCGAGCGCTGGAAGCAAGAGGAGGTAAGAGGAGCAACAGACACACACACC 956
QY 904 CCGACACACATTTAGCCCATTTACAGAGCAGCAGCCCTGCTCCCGAGGCCCAAGACC 963
DB 957 CCGACACACATTTAGCCCATTTACAGAGCAGCAGCCCTGCTCCCGAGGCCCAAGACC 1016
QY 964 AACCTGGGCGAGCGGGGAGAGCAGCCCGCTGTGAACCTTCAAAAGAGAGCTGCC 1023
DB 1017 AACCTGGGCGAGCGGGGAGAGCAGCCCGCTGTGAACCTTCAAAAGAGAGCTGCC 1076
QY 1024 GACTCTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1083
DB 1077 GACTCTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1136
QY 1084 TGCAGCGGCGATCTCAAGAGAGATGTTGCCAAGAACAGCGCCCTGAGCCCTTCC 1143
DB 1137 TGCAGCGGCGATCTCAAGAGAGATGTTGCCAAGAACAGCGCCCTGAGCCCTTCC 1196
QY 1144 TACAGCCTGTGAGCTGAGAGCAGCTGGGCTTACAGACTACTGTGATCATCAAGAC 1203
DB 1197 TACAGCCTGTGAGCTGAGAGCAGCTGGGCTTACAGACTACTGTGATCATCAAGAC 1256
QY 1204 CCCATGACATGAGCAACCAATCAAGCTAACTGAGAGCCCGTGAAGAGCTGATGCTCAG 1263
DB 1257 CCCATGACATGAGCAACCAATCAAGCTAACTGAGAGCCCGTGAAGAGCTGATGCTCAG 1316
QY 1264 GAGTTTGTGCTGACGCTCCGATGATGATGATGATGATGATGATGATGATGATGAT 1323
DB 1317 GAGTTTGTGCTGACGCTCCGATGATGATGATGATGATGATGATGATGATGATGAT 1376
QY 1324 CATGAGTGTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1383
DB 1377 CATGAGTGTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1436
QY 1384 ATGCGGAGAGGAGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1443
DB 1437 ATGCGGAGAGGAGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1496
QY 1444 ACCAAGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1503
DB 1497 ACCAAGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1556
QY 1504 GACAGTTTCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1563
DB 1557 GACAGTTTCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1616
QY 1564 CAGCTCAAAAGCGGTGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1623
DB 1617 CAGCTCAAAAGCGGTGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1676
QY 1624 AAGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1663
DB 1677 AAGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1716

```

RESULT 2  
BC030158 1743 bp mRNA linear HNC 20-MAY-2002  
LOCUS Homo sapiens, similar to bromodomain-containing 4, clone  
DEFINITION IMAGE:5086929, mRNA.  
ACCESSION BC030158



KEYWORDS	REMARK
REVISIONS	
SOURCE	BC030158.1 GI:20988807
ORGANISM	Homo sapiens.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
TITLE	1 (bases 1 to 1743)
AUTHORS	Strausberg, R.
JOURNAL	Direct Submission
COMMENT	Submitted (07-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
COMMENT	NIH-MGC Project URL: <a href="http://mgc.ncl.nih.gov">http://mgc.ncl.nih.gov</a>
COMMENT	Contact: MGC help desk
COMMENT	Email: <a href="mailto:cgaps-remail.nih.gov">cgaps-remail.nih.gov</a>
COMMENT	Tissue Procurement: CGAP (Stanford)
COMMENT	cDNA Library Preparation: Rubin Laboratory
COMMENT	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ULNL)
COMMENT	DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),
COMMENT	Gaithersburg, Maryland;
COMMENT	Web site: <a href="http://www.nisc.nih.gov/">http://www.nisc.nih.gov/</a>
COMMENT	Contact: <a href="mailto:nisc.mgc@nih.gov">nisc.mgc@nih.gov</a>
COMMENT	Blakesley, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bonfield, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Grant, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Latic, P., Legaspi, R., Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Startipop, S., Thomas, P.J., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H., and Green, E.D.
FEATURES	
source	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ULNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Series: IRAL Plate: 39 Row: n Column: 1 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 7657217 This clone has the following problem: frame shifted. Location/Qualifiers 1..1743 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:5086929" /tissue_type="Liver, hepatocellular carcinoma" /clone_id="NIH_MGC_100" /lab_host="DH10B" /note="Vector: pOTB7"
BASE COUNT	487 a 356 c 437 g 263 t
ORIGIN	
Query Match	70.4%; Score 1640; DB 11; Length 1743;
Best Local Similarity	99.4%; Pred. No. 1,9e-284;
Matches 1646;	Conservative 0; Mismatches 10; Indels 0; Gaps 0;
QY	4 GGGATCACTAGCATGTCTGCGGAGAGCGCCCTGGGAGAGATTGAGAAATCTGCCAGTA 63
Db	
QY	87 GGGATCACTAGCATGTCTGCGGAGAGCGCCCTGGGAGAGATTGAGAAATCTGCCAGTA 146
Db	
QY	64 ATGGGGGATGACATGGAACCTTCCCAATGTCTACACACAGGCGCCCAAGCCCAAG 123
Db	
QY	147 ATGGGGGATGACATGGAACCTTCCCAATGTCTACACACAGGCGCCCAAGCCCAAG 206
Db	
QY	124 CCAGCACAAGCAGACGACCAACCCCGCGCCCGCAGAGACCTCCAAACCTTAACAAGCC 183
Db	
QY	207 CCAGCACAAGCAGACGACCAACCCCGCGCCCGCAGAGACCTCCAAACCTTAACAAGCC 266
Db	
QY	184 AAGAGCGAGACCAACCACTGCAATACCTGCTCAGAGTGTGTTCAAGACACTATGAAA 243
Db	
QY	267 AAGAGCGAGACCAACCACTGCAATACCTGCTCAGAGTGTGTTCAAGACACTATGAAA 326
Db	
QY	244 CACCAATTGATGAGCTTTCCAGACAGCTGTGTGATGATGCGTCAAGCTCAACCTCCGTAT 303
Db	

D	327	CACAGTTTGGATGGCCTTTCAGACGACCGCTGTGGATGCGGTCAAGCTGAACCTCCCTCAT	386
O	304	TACTATAAGATCATTTAAAGCCCTATGGATATGGGAACATAAAGAGCGCTTGGAAAC	363
D	387	TACTATAAGATCATTTAAAGCGCCATATGGATATGGGAACATAAAGAGCGCTTGGAAAC	446
O	384	AACATTAACCTGGAAATGCTCAGGAATGTATCCAGACTTCAACACTATGTTTACAAATGT	423
D	447	AACATTAACCTGGAAATGCTCAGGAATGTATCCAGACTTCAACACTATGTTTACAAATGT	506
O	424	TACTATCAACAACGCTGGAGATGACATATGCTTAAATGGACGAAGCTCTGGAAAAAGTC	483
D	507	TACTATCAACAACGCTGGAGATGACATATGCTTAAATGGACGAAGCTCTGGAAAAAGTC	566
O	484	TTCTTGCAAAAAATTAATGAGCTACCCACAGAAAGAAACCGAGATCATGATATGATCCAGCA	543
D	567	TTCTTGCAAAAAATTAATGAGCTACCCACAGAAAGAAACCGAGATCATGATATGATCCAGCA	626
O	544	AAAGGAAGAGACGTGGGAGAAAGAAACAGGACACGCAAAACCTTGGCTTTCCACGTA	603
D	627	AAAGGAAGAGACGTGGGAGAAAGAAACAGGACACGCAAAACCTTGGCTTTCCACGTA	686
O	604	CCAAACACAACTCAAGATGACATCCCTCCGACACACCAGACCCCTCAGCGCAATCTCTCT	663
D	687	CCAAACACAACTCAAGATGACATCCCTCCGACACACCAGACCCCTCAGCGCAATCTCTCT	746
O	664	CCCTGAGAGGCGCACGCTCTCACCCCTTCCCTGCGGTCACCCGAGACTCATCTGTCAGAC	723
D	747	CCCTGAGAGGCGCACGCTCTCACCCCTTCCCTGCGGTCACCCGAGACTCATCTGTCAGAC	806
O	724	CCCTGCATGACATGTTGTCCTCCCGACCGACTGACAGCGCCCCCGCAGTGGCCCCCGAG	783
D	807	CCCTGCATGACATGTTGTCCTCCCGACCGACTGACAGCGCCCCCGCAGTGGCCCCCGAG	866
O	784	CCACAACCCCAACCCCGCTCCAGCTCCCCGACCCGCTACAGACCCCAACCCATATTCGG	843
D	867	CCACAACCCCAACCCCGCTCCAGCTCCCCGACCCGCTACAGACCCCAACCCATATTCGG	926
O	844	GCCACCCCGACGCTGTGAAGACAAAGAGAGAGTAAAGAGAAAGACAGACACCCACCC	903
D	927	GCCACCCCGACGCTGTGAAGACAAAGAGAGAGTAAAGAGAAAGACAGACACCCACCC	986
O	904	CCCAACACCATTTGACCCCATTTCAAGAGCCACCCCTCGCTGCCCGGAGGCCAAGACACC	963
D	987	CCCAACACCATTTGACCCCATTTCAAGAGCCACCCCTCGCTGCCCGGAGGCCAAGACACC	1046
O	964	AAAGTGGGGCCAGCGCGCGGAGAGACAGCCGCGCTGTAAACCTTCAAAAGAGACGTGCC	1023
D	1047	AAAGTGGGGCCAGCGCGCGGAGAGACAGCCGCGCTGTAAACCTTCAAAAGAGACGTGCC	1106
O	1024	GACTCTCAGCAGCACCCAGACAGAGAAAGAGAGCAAGTCTCGAGAGAGCTCAAGTGC	1083
D	1107	GACTCTCAGCAGCACCCAGACAGAGAAAGAGAGCAAGTCTCGAGAGAGCTCAAGTGC	1166
O	1084	TGCAGCGGACATCTCAAGAGATGTTTGCACAAGACACGCGCCTAAGCGCTGGCCCTTC	1143
D	1167	TGCAGCGGACATCTCAAGAGATGTTTGCACAAGACACGCGCCTAAGCGCTGGCCCTTC	1226
O	1144	TACAAGCCTGTGACGTGGAGGACACTGGGCTTACACGACTACTGTGACATCATCAAGAC	1203
D	1227	TACAAGCCTGTGACGTGGAGGACACTGGGCTTACACGACTACTGTGACATCATCAAGAC	1286
O	1204	CCCATGGAACATGAGCAACATCAAGTCTAAACGTGAAGGACCGGATGACCTGATGCTCAG	1263
D	1287	CCCATGGAACATGAGCAACATCAAGTCTAAACGTGAAGGACCGGATGACCTGATGCTCAG	1346
O	1264	GAGTTTGGTGTGACGTCGATGATGATGTTCTCAACTGCTATATAGTAAACCTCTGTAC	1323
D	1347	GAGTTTGGTGTGACGTCGATGATGATGTTCTCAACTGCTATATAGTAAACCTCTGTAC	1406
O	1324	CATGAGTGTGTGACGTCGACCGCGCAAGCTCCAGAGATGTGTGAAATGCGCTTGGCAAG	1383
D	1407	CATGAGTGTGTGACGTCGACCGCGCAAGCTCCAGAGATGTGTGAAATGCGCTTGGCAAG	1466

Oy	1384	ATGCCGAGCAGCCCTGAGGAGCGCAGCGTGCGGCTCTCCCGGCACTGGCCCTCC	1443
Db	1467	ATGCCGAGCAGCCTGAGGAGCGCAGCGTGCGGCTCTCTCCCGGCACTGGCCCTCC	1526
Oy	1444	ACCAAAGTGTGTGGCCCCGCCCTCATTCACGACAGCAGCAGCGATAGCTCTCTGGACAGT	1503
Db	1527	ACCAAAGTGTGTGGCCCCGCCCTCATTCACGACAGCAGCAGCGATAGCTCTCTGGACAGT	1586
Oy	1504	GACAGTTTGACATGATGACCTCTGAGGAGGAGAGGCCACGCGCTGGCTGAGCTCCAGGAG	1563
Db	1587	GACAGTTTGACATGATGACCTCTGAGGAGGAGAGGCCACGCGCTGGCTGAGCTCCAGGAG	1646
Oy	1564	CAGCTCAAAAGCCGTACAGAGCAGCGTTGCAAGCCCTCTCAGCCGCCACGACGAAACCA	1623
Db	1647	CAGCTCAAAAGCCGTACAGAGCAGCGTTGCAAGCCCTCTCAGCCGCCACGACGAAACCA	1706
Oy	1624	AAGAAAGAGCAAGACAGCAAGAGAAAAAGAAAAA	1659
Db	1707	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	1742

RESULT 3	
BC011541	
LOCUS	BC011541
DEFINITION	1776 bp mRNA linear HTC 07-AUG-2007
ACCESSION	Mus musculus, Similar to bromodomain-containing 4, clone IMAGE:2648506, mRNA.
VERSION	BC011541
KEYWORDS	BC011541.1 GI:15079397
SOURCE	HTC.
ORGANISM	house mouse. Mus musculus

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Submitted (30-JUL-2001) National Institutes of Health, Mammalian  
Direct Submission  
Strausberg, R.  
1 (phases 1 to 1776)

REMARK	COMMENT
NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>	Contact: MGC help desk

Contact: MGC help desk  
Email: [cgapds-remail.nih.gov](mailto:cgapds-remail.nih.gov)  
Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNI)  
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
Gunnathre, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,  
Yoon, V.S., Kovis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,  
Richards, S., Gibbs, R.A.

Clone distribution: MCC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://image.llnl.gov>  
Series: IRM Plate: 5 Row: j Column: 5  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 10048443  
This clone has the following problem: Incomplete processing.

**FEATURES**  
**source**

```

/organism="Mus musculus"
/db_xref="taxon:10090"
/map="PVB.N"
/clone="IMAGE:2648506"
/tissue_type="Mammary tumor. Metallothionien-TGf alpha
model. 10 month old virgin mouse. Taken by biopsy."
/clone_id="Nci CGAP Mam1"
/lab_host="DHI0B"
/note="Vector: pcMV-SpOrt6"

```

BASE COUNT	545 a	517 c	401 g	313 t
ORIGIN				

Query Match	60.68;	Score 1411;	DB 11;	Length 1776;
Best Local Similarity	90.38;	Pred. No. 2.4e-243;		
Matches 1520;	Conservative	0;	Mismatches 160;	Indels 3;
				Gaps 1;

Oy	4	GGATATCACTGACATGCTGTGGGAGAGAGGGGCGCTTGGAGACAGATTGTGAAATCTGGCAGTA	63
Db	90	GGGATCACTACATGCTGTACGGAGAGCGGCGCTTGGGACAGATTTGGAATCTGGCAGTA	149
Oy	64	ATGCGGGATGGACTAGAAACCTTCCCAATGTCTACAAACAGAGGCCAGGCCCAACCCAG	123
Db	150	ATGCGGGATGGACTAGAAACCTTCCCAAAATGTCTACAAACAGAGGCCAGGCCCAACCCAG	209
Oy	124	CCAGCCAAAGGAGGCAAGCCACCAACCCCGCGCCCGCCAGAACCTCCAAACCTTAACAAGCC	183
Db	210	CCAGCCAAATGAGGCAAGCCACCAACCTCTCCACCCCGAGAACTCCAAACCTTAACAAGCC	269
Oy	184	AAGAGGAGAGACAACCAACTGCAATATCTGTTCAGAGTGGTGTCTCAAGCACTATGAGAA	243
Db	270	AAGAGCAGAGAACCAACCTGCAATATCTGTTCAGAGTGGTGTCTCAAGCACTATGAGAA	329
Oy	244	CACCAATTTGCATGGCTTTTCCAGACGCTGTGGATGCCGTGAAGCTGAACCTCCGTAT	303
Db	330	CACCAATTTGGGTGGCTTTTCCAGACGCGCGTGGATGCCGTGAACCTGAACCTCCGTAT	389
Oy	304	TACTATTAAGATCATTAAGCGCTATGGATATGAGTAAGCAATTAAGAAAGGCTTGGAAAC	363
Db	390	TACTATTAAGATTTATTAAGACCCATATGATATGGAGACATTAAGAAAGGCTTGGAAAC	449
Oy	364	AACATTAATCTGGAATGTCTCAGGAATGTATCCAGAGACTTCAACAGTATGTTTCAAAATGT	423
Db	450	AACATTAATCTGGAATGTCTCAGGAATGTATCCAGAGACTTCAACAGTATGTTTCAAAATGT	509
Oy	424	TGACATTCACAAAGCGCTGGAGATGACATAGTCTTAATGGCAGGAAGCTCTGGAAAAGCTC	483
Db	510	TGACATTCATACAGCGCTGGAGATGACATAGTCTTAATGGCAGGAAGCTCTGGAAAAGCTC	569
Oy	484	TTCTTGGAAAAAATTAATGAGCTAACCCACAGAAAGAAACCGAATCATGTATGTCCAGCA	543
Db	570	TTCTTGGAAAAAATCATGAACTGCTTACAGAAAGAAACCTGATCATGTATGTCCAGCA	629
Oy	544	AAAGGAGAGAGAGCTGGGAGAGAAAGAAACGAGACGCAAAACCTGGCGTTTCCACGTA	603
Db	630	AAAGGAGAGAGAGAGGAGAGAAAGAAACGAGGAGCAGCAAAACCTGGGTATTCACGTA	689
Oy	604	CCAACACCACTCAAGATGATCTCTCCGAGAGACCACACCCCTCAGCGGAA---TCTT	660
Db	690	CCAACACCACTCAGATCTCAACTTCTCCGAGAGACCACCGCTCTCAGCAGAACCTCTCT	749
Oy	661	CTCTCTGTGACAGGCAAGCCTCAACCCCTTCCGTGCCGTACCCCGGAGCTCATGTGTCAG	720
Db	750	CCACCTTGTGAGGCGCAACATCTACCCCTTCCGTGCTGTCACCCAGACCTCATTTGCCAG	809
Oy	721	ACCCCTGTCAAGACAGTGGTGGCTCCCGACGACATGACAGGCGCCCGCAAGTGGCCCCC	780
Db	810	CTCTCTGTCAAGACATGTGGTGGCTCCCGACGACATTCAGACCTCTTCAACCGGTACCCCC	869
Oy	781	CAGCCACAAACCCCAACCCGCTCCAGCTCCACGACCCGTACAGAGCCACCAACCCATATTC	840
Db	870	CAGCCACAAACCCCAACCCGCTCCAGCTCCAGAGCCGTGCAAGATACCCCGCCATATTT	929
Oy	841	GGCGGCAACCCCAACAGCTGTGAAGACAAAGAAAGGAGTGAAGAGAAAGCAGACACACC	900
Db	930	GGCGGCAACCCCAACCCGCTGTGAAGACAAAGAAAGGAGTGAAGAGAAAGCAGATATCCACC	989
Oy	901	ACCCCAACCACTTGAACCCATTTACAGAGCCACCTCCGCTCCCGGAGGCCAAAGACC	960
Db	990	ACCCCTTACCACTTGAACCCCAATTTCAATGACCAACCTTCACTGGCCCCAGAGGCCAAAGACC	1049
Oy	961	ACCAAGCTGGGGCAGCGGGGAGACAGCCGCGCTGTGTAAACCTTCAAAGAAAGACCTGT	1020

Db 1050 GCCAAGCTGGTCTCGGGGAGAGACGACGACCTGTGAAGCCTCCAAAGAGAGATGA 1109  
QY 1021 CCCGACTCTCAGCAGACCCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
Db 1110 CCGGAGTCACAGCAGACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1169  
QY 1081 TGTGTCAGCGGCATCTTCAGAGAGATGTTTCCAAAGAGAGAGAGAGAGAGAGAGAG 1140  
Db 1170 TGTGTCAGCGGCATCTTCAGAGAGATGTTTCCAAAGAGAGAGAGAGAGAGAGAGAG 1229  
QY 1141 TTCTACAGACCTGTGAG 1200  
Db 1230 TTCTACAGACCTGTGAG 1289  
QY 1201 CACCCCATGAG 1260  
Db 1290 CATCCCATGAG 1349  
QY 1261 CAGAGAGTTGGTGTGAG 1320  
Db 1350 CAGAGAGTTGGTGTGAG 1409  
QY 1321 GACCATGAG 1380  
Db 1410 GACCATGAG 1469  
QY 1381 AAGATGCGCGAG 1440  
Db 1470 AAGATGCGCGAG 1529  
QY 1441 CCCACCAAGTTGTGCGCCCGCTCATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500  
Db 1530 CCGCAAGAGGAG 1589  
QY 1501 AGTACAGAGTTGAG 1560  
Db 1590 AGCAGAGAGTTGAG 1649  
QY 1561 GAGCAGCTCAAG 1620  
Db 1650 GAGCAGCTCAAG 1709  
QY 1621 CCAAG 1680  
Db 1710 CCAAG 1769  
QY 1681 GAA 1683  
Db 1770 AAA 1772

RESULT 4  
B0651563 987 bp mRNA linear EST 15-JUL-2002  
LOCUS B0651563  
DEFINITION AGENCOURT\_8302486 NIH\_MGC\_100 Homo sapiens cDNA clone IMAGE:6271145  
5', mRNA sequence.  
ACCESSION B0651563  
VERSION B0651563.1 GI:21775735  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 987)  
NIH-MGC <http://mgi.nci.nih.gov/>  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-f@mail.nih.gov](mailto:cgaps-f@mail.nih.gov)  
Tissue Procurement: CGAP (Stanford)  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LICM2447 row: 9 column: 18  
High quality sequence start: 2  
High quality sequence stop: 653.  
Location/Qualifiers  
1. 987  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_id="IMAGE:6271145"  
/issue\_type="hepatocellular carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: liver; Vector: pORF7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming, directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACAG(6). Size selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

BASE COUNT 257 a 315 c 267 g 136 t 12 others  
ORIGIN

Query Match 36.8%; Score 856; DB 14; Length 987;  
Best Local Similarity 95.4%; Pred. No. 9,1e-144;  
Matches 885; Conservative 0; Mismatches 42; Indels 1; Gaps 1;

QY 759 GAGCCCCCGCAATGCCCCCGCCAGCAGACCCAGCCGCTCCAGCTCCAGCCCGT 818  
Db 5 GAGCGCTGGGGGCTGGTGGCCCGCCAGCCAGCAACCCAGCCGCTCCAGCTCCAGCCCGT 64

QY 819 ACAGAGCCACCACCCATCATCTGGGGCCAGCCAGCCAGCTGTGAAGCAAGAGAGAGT 878  
Db 65 ACAGAGCCACCACCCATCATCTGGGGCCAGCCAGCCAGCTGTGAAGCAAGAGAGAGT 124

QY 879 GAAG 938  
Db 125 GAAG 184

QY 939 GCTGCCCCCGAGCCAGAGACCAAGCTGGGCGGCGGAGAGAGAGAGAGAGAGAGAG 998  
Db 185 GCTGCCCCCGAGCCAGAGACCAAGCTGGGCGGCGGAGAGAGAGAGAGAGAGAGAG 244

QY 999 GAAACCTCCAAAG 1058  
Db 245 GAAACCTCCAAAG 304

QY 1059 CAAAGTCTCGAG 1118  
Db 305 CAAAGTCTCGAG 364

QY 1119 GCAGCGCGCTAGCGCTGGCCCTTCTACAAAGCTGTGGAGGTGGAGGAGGAGGAGGAG 1178  
Db 365 GCAGCGCGCTAGCGCTGGCCCTTCTACAAAGCTGTGGAGGTGGAGGAGGAGGAGGAG 424

QY 1179 CGACTACTGTGACATCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1238  
Db 425 CGACTACTGTGACATCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 484

QY 1239 GGCCCGTGTACCGTATGATGAG 1298  
Db 485 GGCCCGTGTACCGTATGATGAG 544

QY 1299 CTGCTATAGTACAAACCTCTGACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1358  
Db 545 CTGCTATAGTACAAACCTCTGACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 604

QY 1359 TGTGTTGGAATGGGCTTGGCCAAAGATGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1418  
Db 605 TGTGTTGGAATGGGCTTGGCCAAAGATGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 664

QY	1419	GTCTCCCTCCCGGAGTGGCCCTCTCCACCAAGTTGTGGCCCGGCTCATCCAGGACAG	1478
Db	665	GTCTCTCCCGGAGTGGCCCTCTCCACCAAGTTGTGGCCCGGCTCATCCAGGACAG	724
QY	1479	CAGCAGCGATATAGTCTCTTGGGACATGTGACAGTTTCGACTGATCTGTAGAGGAGCGGAC	1538
Db	725	CAGCAGCGATATAGTCTCTTGGGACATGTGACAGTTTCGACTGATCTGTAGAGGAGCGGAC	784
QY	1539	CCAGCGGTGTGGTGAGTCCAGAGGACGACCTCAACCCGTGCACGAGCAGCTTGACGCCCT	1598
Db	785	CCAGCGGTGTGGTGAGTCCAGAGGACGACCTCAACCCGTGCACGAGCAGCTTGACGCCCT	843
QY	1599	CTCTCAGCCCCCAGCAAAACAAACCAAGAAAAAGGAGAAAGCAAGAGAAAAA	1658
Db	844	CTCTCAGCCCCCAGCAAAACAAACCCCAAAAAAAAAAAAAAAAAAGGANNNNNNNANNA	903
QY	1659	AGAAAAGCAAAAAGAAAAGAGAGATG	1686
Db	904	GGAAAAGCAAAAAGAAAAGAGAGATG	931

RESULT 5					
B0646784					
LOCUS	B0646784	951 bp	mRNA	linear	EST 15-JUL-2002
DEFINITION	AGENCOURT_8298482 NIH_MGC_100 Homo sapiens CDNA clone IMAGE:6270294				
ACCESSION	5', mRNA sequence.				
	B0646784				

ACCESSION	BQ646784
VERSION	BQ646784.1
REVISION	GI:21770956

KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 951)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.

Email: [cgabos-remail.nih.gov](mailto:cgabos-remail.nih.gov)  
Tissue Procurement: CGAP (Stanford)  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
Plate: LICM2445 row: d column: 07  
High quality sequence stop: 636.

FEATURES

SOURCE

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6270294"
/clone_1db="NIH_MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pOTB1; Site.1: XhoI; Site.2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCGGAGG(c). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."

```

BASE COUNT	271 a	305 c	226 g	148 t	1 others
ORIGIN					

Query Match	33.8%	Score 787	DB 14	Length 951
Best Local Similarity	96.5%	Pred. No. 2.2e-111		
Matches 836	Conservative 0	Mismatches 26	Indels 4	Gaps 3

QY	4	GGGATCTACATCACTGCTGCGGAGAGCGCCCTGGGAGCAATTTGAGAAATCTGCACTA	63
Db	77	GGGATCTACATCACTGCTGCGGAGAGCGCCCTGGGAGCAATTTGAGAAATCTGCACTA	136
QY	64	ATGGGGATGCACTAGAAACTTCCCAATATGCTACACACAGGCCAGGCCCAACCCAG	123
Db	137	ATGGGGATGCACTAGAAACTTCCCAATATGCTACACACAGGCCAGGCCCAACCCAG	196
QY	124	CCAGCCACAGGAGCGAGCACACACCCCGCCCGCCAGAGACTCCCAACCTTACAAAGCC	183
Db	197	CCAGCCACAGGAGCGAGCACACACCCCGCCCGCCAGAGACTCCCAACCTTACAAAGCC	256
QY	184	AAGAGCGAGACCAACCAACTGCATACCTGCTCAGAGTGTGCTCAAGACACTATGAAA	243
Db	257	AAGAGCGAGACCAACCAACTGCATACCTGCTCAGAGTGTGCTCAAGACACTATGAAA	316
QY	244	CACCAAGTTGCATGAGCCCTTCCAGCAGCGCTGTGATGTCGCCGTCAAGCTGACCTCCGTAT	303
Db	317	CACCAAGTTGCATGAGCCCTTCCAGCAGCGCTGTGATGTCGCCGTCAAGCTGACCTCCGTAT	376
QY	304	TACTATAGATCATTTAAACCCCTATGGATATGGGAACATTAAGAAAGCGCTTGGAAAC	363
Db	377	TACTATAGATCATTTAAACCCCTATGGATATGGGAACATTAAGAAAGCGCTTGGAAAC	436
QY	364	AACTATTACTGGAATGCTCAGGAATGTATGCCAGACTTCAACACTATGTTTACAAATTTGT	423
Db	437	AACTATTACTGGAATGCTCAGGAATGTATGCCAGACTTCAACACTATGTTTACAAATTTGT	496
QY	424	TACATCTCAACAAGCCCTGGAGATGACATAGCTTTAATGGGAGAAGCTCTGAAAAAGCTC	483
Db	497	TACATCTCAACAAGCCCTGGAGATGACATAGCTTTAATGGGAGAAGCTCTGAAAAAGCTC	556
QY	484	TTCTTGGCAAAAAATTAATAGACTACCCACAGAAAGAACCGAGATCATGATAGTCCAGGCA	543
Db	557	TTCTTGGCAAAAAATTAATAGACTACCCACAGAAAGAACCGAGATCATGATAGTCCAGGCA	616
QY	544	AAAGGAACAGAGACGTGGGAGGAAAGAAACAGGAGACAGCAAAACCTGGCTTTCCAGGTA	603
Db	617	AAAGGAACAGAGACGTGGGAGGAAAGAAACAGGAGACAGCAAAACCTGGCTTTCCAGGTA	676
QY	604	CCAAAACACAACTCAAGCATCTCCTCCGAGAGACCCAGACCCCTCAGCCCAATCTT -CC	662
Db	677	CCAAAACACAACTCAAGCATCTCCTCCGAGAGACCCAGACCCCTCAGCCCAATCTCTCC	736
QY	663	TCTCTGTGAGGCGCAAGCCTCAACCCCTTCCCTGCGCGTCAACCCCGAGACTCATCTGTCCAGAC	722
Db	737	TCTCTGTGAGGCGCAAGCCTCAACCCCTTCCCTGCGCGTCAACCCCGAGACTCATCTGTCCAGAC	796
QY	723	CCCTGTCAATGACAGTGGTGCCTCCGACAGCACTGAGAGCG - -CCGCCGCGAGTGGCCCCC	780
Db	797	CCCTGTCAATGACAGTGGTGCCTCCGACAGCACTGAGAGCG - -CCGCCGCGAGTGGCCCCC	856
QY	781	CAGCCACAAACCCCAACCCGCTTCCAGCTTCCCGACGCCCGGTACAGAGCCACCCACCATCATCTC	840
Db	857	CAGCCACAAACCCCAACCCGCTTCCAGGCTTCCAGCCCGGAAAGAGCAGCCACCATCATTTTC	916
QY	841	GC -GGCCACCCCAAGCCTGTGAAGA	865
Db	917	GCGGCCACCCCAAGCCTGGGAAAA	942

RESULT 6	
LOCUS	BQ219777
DEFINITION	BQ219777 893 bp mRNA linear EST 02-MAY-2002
	ACENCCOURT_7644864 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:6020711
	5', mRNA sequence.

ACCESSION	BQ219777	
VERSION	BQ219777.1	GI:20401177

KEYWORDS	EST.
SOURCE	human.

ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 893)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>,  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: The I.M.A.G.E. Consortium (LNLN)  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNLN at:  
<http://image.llnl.gov>  
Plate: LLM13225 Row: d Column: 24  
High quality sequence stop: 675.  
Location/Qualifiers  
1..893  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6020711"  
/clone\_lib="NIH\_MGC\_70"  
/tissue\_type="epithelioid carcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dt.  
Average insert size 1.1 kb. Library constructed by Life  
Technologies."

BASE COUNT 259 a 298 c 199 g 137 t

ORIGIN

Query Match 33.4%; Score 779; DB 14; Length 893;  
Best Local Similarity 95.3%; Pred. No. 6e-130;  
Matches 826; Conservative 0; Mismatches 35; Indels 6; Gaps 2;

QY 200 AACGCAATACCTGCTGAGAGTGGTGCACAGACATATGGAACACAGTTGCAATGGC 259  
Db 1 AACTGCAATACCTGCTGAGAGTGGTGCACAGACATATGGAACACAGTTGCAATGGC 60  
QY 260 CTTTCAGCAGCCTGTGATGCTGCTCAAGCTGCAAGCTCCCTGATTACTATTAAGATCATTA 319  
Db 61 CTTTCAGCAGCCTGTGATGCTGCTCAAGCTGCAAGCTCCCTGATTACTATTAAGATCATTA 120  
QY 320 AAAGCCTATGATATGGAACATTAAGAAAGCCTTGGAACAACTATTACTGGAATG 379  
Db 121 AAAGCCTATGATATGGAACATTAAGAAAGCCTTGGAACAACTATTACTGGAATG 180  
QY 380 CTCGAGATGATCAGAGACTTCAACACTATGTTACAAATTTTACATCTACAAACAGC 439  
Db 181 CTCGAGATGATCAGAGACTTCAACACTATGTTACAAATTTTACATCTACAAACAGC 240  
QY 440 CTGAGATGATCAGAGACTTCAACACTATGTTACAAATTTTACATCTACAAACAGC 499  
Db 241 CTGAGATGATCAGAGACTTCAACACTATGTTACAAATTTTACATCTACAAACAGC 300  
QY 500 ATGAGCTATCCACAGAAAGAAACGAGATCATGATGTCAGGCAAAAGAAAGAGAGAGCTG 559  
Db 301 ATGAGCTATCCACAGAAAGAAACGAGATCATGATGTCAGGCAAAAGAAAGAGAGAGCTG 360  
QY 560 GGAAGAAAGAAAGAGAGAGCAAAACCTGGCGTTTCCAGGTCACAAACCAACTCAAG 619  
Db 361 GGAAGAAAGAAAGAGAGAGCAAAACCTGGCGTTTCCAGGTCACAAACCAACTCAAG 420  
QY 620 CATGACTCTCCGACAGACCCAGACCCCTCAGCGAAATCTCTCTGTGAGGAGCCAGC 679  
Db 421 CATGACTCTCCGACAGACCCAGACCCCTCAGCGAAATCTCTCTGTGAGGAGCCAGC 480  
QY 680 CTCACCCCTTCTGCGCTGCTACCCCGAGACTCATGTCACAGACCCCTGTATGACAGTGG 739  
Db 481 CTCACCCCTTCTGCGCTGCTACCCCGAGACTCATGTCACAGACCCCTGTATGACAGTGG 540  
QY 740 TGGCTCCCAAGCCAGAGAGCCCGGAGAGTGGCCCGCCAGCCCAACCCCGACCGG 799  
Db 541 TGGCTCCCAAGCCAGAGAGCCCGGAGAGTGGCCCGCCAGCCCAACCCCGACCGG 600

QY 800 CTCAGCTCCCCAGCCCTTACAGAGCCACCCACCATCATGCGGCGCACCCAGAGCCTG 859  
Db 601 CTCAGCTCCCCAGCCCTTACAGAGCCACCCACCATCATGCGGCGCACCCAGAGCCTG 660  
QY 860 TGAAGACAAAGAGAGAGTGAAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 919  
Db 661 TGAAGACAAAGAGAGAGTGAAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720  
QY 920 CCATTACAGAGCCACCCCTGCTGCCCCCGAGCCCAAGACCAAGTGGGCCAGCGCG 979  
Db 721 CCATTACAGAGCCACCCCTGCTGCCCCCGAGCCCAAGACCAAGTGGGCCAGCGCG 780  
QY 980 GGGAGAG---CAGCCGCGCTGTGA---CCTCCAAAGAAAGAGAGTGGCTCTGAGC 1033  
Db 781 GGGAGAGAGAGAGAGCGCGGCGCTGTGA---CCTCCAAAGAAAGAGAGTGGCTCTGAGC 840  
QY 1034 AGCACCAGCAGCAGAGAGAGAGAGCA 1060  
Db 841 AGCACCAGCAGCAGCAGAGAGAGAGAA 867

RESULT 7  
BQ214403 933 bp mRNA linear EST 02-MAY-2002  
LOCUS AGENCOURT\_7589671 NIH\_MGC\_92 Homo sapiens CDNA clone IMAGE:6067786  
DEFINITION 5', mRNA sequence.  
ACCESSION BQ214403  
VERSION BQ214403.1 GI:20395803  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 933)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>,  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNLN at:  
<http://image.llnl.gov>  
Plate: LLM13347 Row: n Column: 11  
High quality sequence stop: 659.  
Location/Qualifiers  
1..933  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6067786"  
/clone\_lib="NIH\_MGC\_92"  
/tissue\_type="embryonal carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: testis; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 2.5 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC Library."

BASE COUNT 288 a 304 c 196 g 145 t

ORIGIN

Query Match 33.1%; Score 770.2; DB 14; Length 933;  
Best Local Similarity 92.3%; Pred. No. 2.3e-128;  
Matches 846; Conservative 0; Mismatches 63; Indels 8; Gaps 3;

QY 25 GAGAGCGCCCTGGGAGAGATGAGAAATCTGCAGTAATGGGGATGGAGTGAAGTACT 84  
Db 13 GCGAGCGCCCTGGGAGAGATGAGAAATCTGCAGTAATGGGGATGGAGTGAAGTACT 72

```
OY 85 TCCCAATGCTCTACACACAGGCGCCAGGCCAACCCAGCCAGCAACGACGACAGC 144
    |||||||
Db 73 TCCCAATGCTCTACACACAGGCGCCAGGCCAACCCAGCCAGCAACGACGACAGC 132
OY 145 AACCCCCCGCCCGCAGAGAGCTTCAACCCCTAACAGGCCAGGAGGACCAACCACTG 204
    |||||||
Db 133 AACCCCCCGCCCGCAGAGAGCTTCAACCCCTAACAGGCCAGGAGGACCAACCACTG 192
OY 205 CAATACCTGCTCAGAGTGGTGTCTCAAGACACTATGGAACACAGTTTGCATGCTTTC 264
    |||||||
Db 193 CAATACCTGCTCAGAGTGGTGTCTCAAGACACTATGGAACACAGTTTGCATGCTTTC 252
OY 265 CACGACCTGTGTGATGCGCTCAAGCTGAACCTCCGATTACTATTAAGATCAATTAACG 324
    |||||||
Db 253 CACGACCTGTGTGATGCGCTCAAGCTGAACCTCCGATTACTATTAAGATCAATTAACG 312
OY 325 CCTATGATATGGAACCAATTAAGAACCGCTTGGAAACCACTATTACTGGAATGCTCAG 384
    |||||||
Db 313 CCTATGATATGGAACCAATTAAGAACCGCTTGGAAACCACTATTACTGGAATGCTCAG 372
OY 385 GAATGATCCAGAGCTTCAACACTATGTTTACAATTTGTACATCTAACCAAGCCTGGA 444
    |||||||
Db 373 GAATGATCCAGAGCTTCAACACTATGTTTACAATTTGTACATCTAACCAAGCCTGGA 432
OY 445 GATGACATATGCTTAAATGGCAGAGCTGTGAAAGCTCTCTTGTGCAAAAAATTAATGAG 504
    |||||||
Db 433 GATGACATATGCTTAAATGGCAGAGCTGTGAAAGCTCTCTTGTGCAAAAAATTAATGAG 492
OY 505 CTACCCACAGAAAGAACCCAGATCATATGATGTCAGGCAAAAGGAAGAGAGCTGGGAG 564
    |||||||
Db 493 CTACCCACAGAAAGAACCCAGATCATATGATGTCAGGCAAAAGGAAGAGAGCTGGGAG 552
OY 565 AAAGAAACAGGAGACAGCAAAACCTGGCGTTTCCACGCTACCAAAACCACTCAAGCATCG 624
    |||||||
Db 553 AAAGAAACAGGAGACAGCAAAACCTGGCGTTTCCACGCTACCAAAACCACTCAAGCATCG 612
OY 625 ACTGCTCCGAGAGCCAGACCCCTCAGCCGATCTCTCTCTGAGAGGCGCAGCCCTCAG 684
    |||||||
Db 613 ACTGCTCCGAGAGCCAGACCCCTCAGCCGATCTCTCTCTGAGAGGCGCAGCCCTCAG 672
OY 685 CCGTTCCCTGCGCTCAACCCCGGACCTCATCTGTCAGACCCCTGTCATGACATGTCCT 744
    |||||||
Db 673 CCGTTCCCTGCGCTCAACCCCGGACCTCATCTGTCAGACCCCTGTCATGACATGTCCT 732
OY 745 CCGCAGCCTGCTGACAGCGCCCGCCAG--TGCCTCCCGCAGCCCAACCCCGCCTC 802
    |||||||
Db 733 CCGCAGCCTGCTGACAGCGCCCGCCAGGTCGCTCCCAAGCACAACCCCGCCTCAG 792
OY 803 C-AGCTCCCGACGCGGTACAGAGCAGCCACCC-----ATCATCGGGGGCCAGCCAGC 856
    |||||||
Db 793 CAAGCTCCCAAGGCCCAACAGAGCCACCCCAACCCATCATTTCCGGGGCCACCCCGCAAGGC 852
OY 857 CTGTGAAGACAAAGAGGAGTGAAGAGAGCAACACACCCCGCAGCCAGCATTTG 916
    |||||||
Db 853 CTGTGGGAAGAGCAAAAAAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 912
OY 917 ACCCCATTACAGAGCCA 933
    |||||||
Db 913 TCCCGCCACCAACCCA 929
```

```
RESULT 8
BO653187 975 bp mRNA linear EST 15-JUL-2002
LOCUS DEFINITION AGENCODRT_8341467 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6267841
5', mRNA sequence.
ACCESSION BO653187
VERSION BO653187.1 GI:21777359
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE 1 (bases 1 to 975)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LLM2438 row: n column: 02
High quality sequence stop: 656.
Location/Qualifiers
1. 975
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1ib="NIH_MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GCCACGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-CDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH-MGC
Library."
BASE COUNT 275 a 304 c 246 g 145 t 5 others
ORIGIN
Query Match 32.5%; Score 756.6; DB 14; Length 975;
Best Local Similarity 93.7%; Pred. No. 6.2e-126;
Matches 841; Conservative 0; Mismatches 49; Indels 8; Gaps 5;
OY 4 GGGATCACTAGCATGTCTGGGAGAGCGGCCCTGGGAGAGATTGGAATCTGCCAGTA 63
    |||||||
Db 77 GGGATCACTAGCATGTCTGGGAGAGCGGCCCTGGGAGAGATTGGAATCTGCCAGTA 136
OY 64 ATGGGGATGAGCTAGAAAACCTTCCAAATGCTTCAACACAGGCGCCCAACCCAG 123
    |||||||
Db 137 ATGGGGATGAGCTAGAAAACCTTCCAAATGCTTCAACACAGGCGCCCAACCCAG 196
OY 124 CCAAGCAAGCGACGACGACCAACCCCGCCCGCCAGAGACCTCAACCTTAACAAGCC 183
    |||||||
Db 197 CCAAGCAAGCGACGACGACCAACCCCGCCCGCCAGAGACCTCAACCTTAACAAGCC 256
OY 184 AAGAGCAGACCAACCACTGCAATACCTGCTCAGAGTGTGCTCAAGACATATGAAA 243
    |||||||
Db 257 AAGAGCAGACCAACCACTGCAATACCTGCTCAGAGTGTGCTCAAGACATATGAAA 316
OY 244 CACCACTTTCATAGGCTTTCCAGAGCTGTGATGCTGCTCAAGCTGAACCTCCTGAT 303
    |||||||
Db 317 CACCACTTTCATAGGCTTTCCAGAGCTGTGATGCTGCTCAAGCTGAACCTCCTGAT 376
OY 304 TACTATAGATCACTTAAAGCGCTATGATATGGAACAATAAGAAAGGCTTGAAGAAC 363
    |||||||
Db 377 TACTATAGATCACTTAAAGCGCTATGATATGGAACAATAAGAAAGGCTTGAAGAAC 436
OY 364 AACTATTACTGAAATGCTCAGAAATATCAAGAGCTTCAACACTATGTTTACAATTTGT 423
    |||||||
Db 437 AACTATTACTGAAATGCTCAGAAATATCAAGAGCTTCAACACTATGTTTACAATTTGT 496
OY 424 TACATCTACAACAAGCCTGAGATGACATAGTCTTAATGAGCAGAACTCTGAAAAGCTC 483
    |||||||
Db 497 TACATCTACAACAAGCCTGAGATGACATAGTCTTAATGAGCAGAACTCTGAAAAGCTC 556
    |||||||
OY 484 TTCTTTCAAAAAATAATAGCTACCCACAGAAAGAAAGGAGATCTGATAGTCCAGGCA 543
    |||||||
```



Query Match	32.3%	Score 752.4	DB 14	Length 925
Best Local Similarity	97.0%	Pred. No. 3.5e-125		
Matches 776	Conservative 0	Mismatches 23	Indels 1	Gaps 1
QY 4	GGGATCTAGCATGTCCTGGGAGGCGCCCTGGGAGAGATTGAGAAATCTGCGACTA	63		
Db 77	GGGATCTAGCATGTCCTGGGAGGCGCCCTGGGAGAGATTGAGAAATCTGCGACTA	136		
QY 64	ATGGGGATGACACTAGAAACTTCCCAATATGCTACACACAGGCCAGGCCAACCCGAG	123		
Db 137	ATGGGGATGACACTAGAAACTTCCCAATATGCTACACACAGGCCAGGCCAACCCGAG	196		
QY 124	CCAGCCAAACGACGACGACGACCAACCCCGCCCGCCAGAGACCTCCACCTAACAGCCC	183		
Db 197	CCAGCCAAACGACGACGACGACCAACCCCGCCCGCCAGAGACCTCCACCTAACAGCCC	256		
QY 184	AAGAGGACGACCAACCAACTGCAATACCTGCTCAGAGTGTGTCTAAACACTATGGAAA	243		
Db 257	AAGAGGACGACCAACCAACTGCAATACCTGCTCAGAGTGTGTCTAAACACTATGGAAA	316		
QY 244	CACCAAGTTTGCATGGCTTTCCAGAGACGCTGGATGGATGGCGTCAAGCTAACCTCCGTAT	303		
Db 317	CACCAAGTTTGCATGGCTTTCCAGAGACGCTGGATGGATGGCGTCAAGCTAACCTCCGTAT	376		
QY 304	TACTATTAAGATCATTTAAACGCTATGATATGGAGAACATTAAGAAAGCGTTGGAAAC	363		
Db 377	TACTATTAAGATCATTTAAACGCTATGATATGGAGAACATTAAGAAAGCGTTGGAAAC	436		
QY 364	AACCTATTACCTGGAATGCTCAGGAATGTATCCAGACTTCAACACTATTTTACAAATGT	423		
Db 437	AACCTATTACCTGGAATGCTCAGGAATGTATCCAGACTTCAACACTATTTTACAAATGT	496		
QY 424	TACATCTCAACAACGCTGGAGATGACATGATGCTTAATGGCAGAAAGCTCTGGAAAGCTC	483		
Db 497	TACATCTCAACAACGCTGGAGATGACATGATGCTTAATGGCAGAAAGCTCTGGAAAGCTC	556		
QY 484	TTCTTGCAAAAAATTAATGAGCTACCCACAGAAAGAACCGAGATCATGATAGTCCAGCA	543		
Db 557	TTCTTGCAAAAAATTAATGAGCTACCCACAGAAAGAACCGAGATCATGATAGTCCAGCA	616		
QY 544	AAAGGAAGGACGCTGGGAGGAAAGAAAACAGGGACAGCAAAACCTGGCGTTTCCAGGTA	603		
Db 617	AAAGGAAGGACGCTGGGAGGAAAGAAAACAGGGACAGCAAAACCTGGCGTTTCCAGGTA	676		
QY 604	CCAAACACAACTCAAGCATGCACTGCTCGGAGACCCGACGCCGAATCTCTCT	663		
Db 677	CCAAACACAACTCAAGCATGCACTGCTCGGAGACCCGACGCCGAATCTCTCTCT	736		
QY 664	CCTGTGCAAGGCGACGCTCACCCCTTCCCTGCGCTGACACCCGAGACTGATGTCAGAGC	723		
Db 737	CCTGTGCAAGGCGACGCTCACCCCTTCCCTGCGCTGACACCCGAGACTGATGTCAGAGC	796		
QY 724	CCTGTGCAATGAC-AGTGTGTCCTCCCGACGACCTGACAGCGCCCGCGCATGTGCCCCCA	782		
Db 797	CCTGTGCAAGAAAGTGTGTCCTCCANNCCCTGTGAGAGCCCGCGCAGGGGCGCCCC	856		
QY 783	GGCAACAACCCCGACCCGCTC 802			
Db 857	CCAACCAACCCCGACCCGCTC 876			
RESULT 10				
LOCUS B0647311	899 bp	mRNA	linear	EST 15-JUL-2002
DEFINITION AGNC00UR1.8350802 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6287170				
ACCESSION B0647311				
VERSION B0647311.1	GI:21771483			
KEYWORDS EST.				
SOURCE human.				
ORGANISM Homo sapiens				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				

REFERENCE 1 (bases 1 to 899)  
AUTHORS NIH-MGC http://mhc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strusberg, Ph.D.  
Email: cgapbs-remail.nih.gov

Tissue Procurement: CGAP (Stanford)  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LICM2489 row: c column: 11  
High quality sequence stop: 680.  
Location/Qualifiers

## FEATURES

source

1. 899  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6287170"  
/clone.lib="NIH-MGC.100"  
/tissue.type="hepatocellular carcinoma, cell line"  
/lab.host="DH10B (phage-resistant)"  
/note="Organ: liver; Vector: pOTB7, Site\_1: XhoI; Site\_2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC Library."

BASE COUNT 257 a 278 c 217 g 144 t 3 others  
ORIGIN

Query Match 31.6%; Score 736.2; DB 14; Length 899;  
Best Local Similarity 95.4%; Pred. No. 2.8e-122;

Matches 768: Conservative 0; Mismatches 35; Indels 2; Gaps 1;

4 GGGATCCTAGCATGTCTCGGAGAGCGCCCTGGGACGAGATTGGAATCTGCCAGTA 63  
77 GGGATCCTAGCATGTCTCGGAGAGCGCCCTGGGACGAGATTGGAATCTGCCAGTA 136  
64 ATGGGGGATGACTAGAAAATCTCCAAATGTCTACAAACAGGCCAGGCCCAACCCAG 123  
137 ATGGGGGATGACTAGAAAATCTCCAAATGTCTACAAACAGGCCAGGCCCAACCCAG 196  
124 CGAGCCAGCGAGCGAGCAACCCCGCCCGCCGAGAGACCTCCAAACCTTAACAGCCC 183  
197 CGAGCCAGCGAGCGAGCAACCCCGCCCGCCGAGAGACCTCCAAACCTTAACAGCCC 256  
184 AAGAGCGAGACCAACCACTGCATACCTGCTCAGAGTGGTCTCAAGACTATGGAA 243  
257 AAGAGCGAGACCAACCACTGCATACCTGCTCAGAGTGGTCTCAAGACTATGGAA 316  
244 CACCACTTGCATGGCCTTCCAGACGCTGGGATGCCGTCAGCTGAACCTCCCTGAT 303  
317 CACCACTTGCATGGCCTTCCAGACGCTGGGATGCCGTCAGCTGAACCTCCCTGAT 376  
304 TACTATAGATCAATAAAGCGCTATGATGATGAGCAATAAAGAGCGCTGGAAAC 363  
377 TACTATAGATCAATAAAGCGCTATGATGATGAGCAATAAAGAGCGCTGGAAAC 436  
364 AACTATTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 423  
437 AACTATTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 496  
424 TACATCTACACAGCGCTGGAGATGACATATGCTTAATGAGCAAGCGCTGGAAAAAGCTC 483  
497 TACATCTACACAGCGCTGGAGATGACATATGCTTAATGAGCAAGCGCTGGAAAAAGCTC 556  
484 TTCTTGCAAAAAATTAATGAGCTACCCACAGAAAGCAAGCAAGCAAGCAAGCAAGCA 543

|||||  
Db 557 TTCTTGCAAAAAATTAATGAGCTACCCACAGAAAGCGAGATCAATGATCCAGGCA 616  
Oy 544 AAGAGAGAGAGAGCTGGAGAGAAAGAAAGAGAGAGCAAAACCTGGCTTCCAGGTA 603  
Db 617 AAGAGAGAGAGAGCTGGAGAGAAAGAAAGAGAGAGCAAAACCTGGCTTCCAGGTA 676  
Oy 604 CCAACACACATCTACAGATGATCTCTCCGAGAGACCCAGACCCCTGAGCAATCTCT 663  
Db 677 CCAACACACATCTACAGATGATCTCTCCGAGAGACCCAGACCCCTGAGCAATCTCT 736  
Oy 664 CCTGTGACAGCCAGCGCTTACCCCTTCCCTGCGCTACACCCGAGCTCATCTCCAGAC 723  
Db 737 CCTGTGACAGCCAGCGCTTACCCCTTCCCTGCGCTACACCCGAGCTCATCTCCAGAC 796  
Oy 724 CCTGTGATG--ACAGTGTGCTCCCGACGACCTGACAGGCGCCCGCAGTGCCCGCC 781  
Db 797 CCTGTATGAAACAGTGTGCTCCCGACGACCTGAGGAGAAAGCGCCCGCAGAGGC 856  
Oy 782 AGCCACACCCCGACCGCTCCAGC 806  
Db 857 CCCCCCAGCAAAAGCCCCCAGC 881

RESULT 11  
B0650713 931 bp mRNA linear EST 15-JUL-2002  
LOCUS AGENCOURT.8419166 NIH-MGC\_100 Homo sapiens CDNA clone IMAGE:6284393  
DEFINITION 5', mRNA sequence.  
ACCESSION B0650713  
VERSION B0650713.1 GI:21774885  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 931)  
AUTHORS NIH-MGC http://mhc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strusberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: CGAP (Stanford)  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LICM2481 row: c column: 18  
High quality sequence stop: 639.  
Location/Qualifiers

## FEATURES

source

1. 931  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6284393"  
/clone.lib="NIH-MGC.100"  
/tissue.type="hepatocellular carcinoma, cell line"  
/lab.host="DH10B (phage-resistant)"  
/note="Organ: liver; Vector: pOTB7, Site\_1: XhoI; Site\_2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC Library."

BASE COUNT 258 a 299 c 219 g 149 t 6 others  
ORIGIN

Query Match 31.6%; Score 735.8; DB 14; Length 931;  
Best Local Similarity 95.8%; Pred. No. 3.4e-122;



Matches 775; Conservative 0; Mismatches 31; Indels 3; Gaps 2;			
QY	4	GGATCACTACATGCTGTGCGAGAGCGCCCTGGAGAGATTTAGAAATCTGCCAGTA	63
Db	77	GGATCACTACATGCTGTGCGAGAGCGCCCTGGAGAGATTTAGAAATCTGCCAGTA	136
QY	64	ATGGGGATGACATAGAACTTCCCAATGTCTACACACAGGCCCGCCCAACCCAG	123
Db	137	ATGGGGATGACATAGAACTTCCCAATGTCTACACACAGGCCCGCCCAACCCAG	196
QY	124	CCAGCCAAAGGAGGACGACACCAACCCCCCGCCCGCCAGAGACCTCCAAACCC	183
Db	197	CCAGCCAAAGGAGGACGACACCAACCCCCCGCCCGCCAGAGACCTCCAAACCC	256
QY	184	AAGAGGAGACCAACCACTACATACCTGTCTAGAGTGGTGTCTCAAGACACTAG	243
Db	257	AAGAGGAGACCAACCACTACATACCTGTCTAGAGTGGTGTCTCAAGACACTAG	316
QY	244	CACCACTTTGATGCGCTTTCCAGACGCTGTGATGCGCTCAAGCTGAACCTCC	303
Db	317	CACCACTTTGATGCGCTTTCCAGACGCTGTGATGCGCTCAAGCTGAACCTCC	376
QY	304	TACTATAGATTCATTAAGAGCCATATGATATGGGAACAATAAAGCCCTTGGAA	363
Db	377	TACTATAGATTCATTAAGAGCCATATGATATGGGAACAATAAAGCCCTTGGAA	436
QY	364	AACATTTACTGATGCTAGAGATGTATCCAGAGCTTCAACACTATGTTTACAAT	423
Db	437	AACATTTACTGATGCTAGAGATGTATCCAGAGCTTCAACACTATGTTTACAAT	496
QY	424	TACATCTACACAAGCCTGAGATGACATAGTCTTAATGCGAGAAGCTTGGAAA	483
Db	497	TACATCTACACAAGCCTGAGATGACATAGTCTTAATGCGAGAAGCTTGGAAA	556
QY	484	TTCTTGGAAAAATTAATGAGCTTACGACAGAAAGAACCGAGATCATGATGTCC	543
Db	557	TTCTTGGAAAAATTAATGAGCTTACGACAGAAAGAACCGAGATCATGATGTCC	616
QY	544	AAAGGAAGAGAGCTGGGAGGAAAGAAACAGAGACACCAAAACCTGGCTTCC	603
Db	617	AAAGGAAGAGAGCTGGGAGGAAAGAAACAGAGACACCAAAACCTGGCTTCC	676
QY	604	CAAAACAACCTCAAGCATGCTCTCCGACAGACCCGCTCAAGCCGAACTCTCT	663
Db	677	CAAAACAACCTCAAGCATGCTCTCCGACAGACCCGCTCAAGCCGAACTCTCT	736
QY	664	CGTGTGAGGCGACGCTCAACCCCTTCCCTGCGCTCAACCCGAGCTCATGTCT	723
Db	737	CGTGTGAGGCGACGCTCAACCCCTTCCCTGCGCTCAACCCGAGCTCATGTCT	796
QY	724	CTGTGATG-ACAGTGTGCTCCCGAG--CACTGAGAGAGCCCGCCAGTGCC	780
Db	797	CTGTGATGAGCAAGTGGGTGCTCCGAGGAGCCCTGAGAGAGGCGCCGCAAT	856
QY	781	CAGCCACACCCCGGCTCCAGCTCC 809	
Db	857	CCCCAGGACACACCCCGGCTCC 885	
RESULT 12			
LOCUS B0652326 919 bp mRNA linear EST 15-JUL-2002			
DEFINITION AGENCOUTRT_8350731 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6287333			
ACCESSION B0652326			
VERSION B0652326.1 GI:21776498			
KEYWORDS EST.			
SOURCE human.			
ORGANISM Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE 1 (bases 1 to 919)			
AUTHORS NIH-MGC <a href="http://mgc.ncl.nih.gov/">http://mgc.ncl.nih.gov/</a> .			

TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)									
JOURNAL	Unpublished (1999)									
COMMENT	Contact: Robert Strausberg, Ph.D. Email: gcgaps-remail.nih.gov Tissue Procurement: CGAP (Stanford) cDNA Library Preparation: Rubin Laboratory DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <a href="http://image.lnl.nih.gov">http://image.lnl.nih.gov</a> Plate: LUCM2489 row: j column: 06 High quality sequence stop: 593. Location/Qualifiers									
FEATURES	1. 919 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:6287333" /clone_11b="NIH.MGC.100" /tissue.type="hepatocellular carcinoma, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: liver; Vector: pOTB1; Site_1: XhoI; Site_2: EcoRI; cDNA made by Oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCGAGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC Library."									
BASE COUNT	259 a	289 c	221 g	149 t	1 others					
ORIGIN										
Query Match	31.5%; Score 733.4; DB 14; Length 919;									
Best Local Similarity	94.5%; Pred. No. 9.1e-122;									
Matches 793; Conservative 0; Mismatches 41; Indels 5; Gaps 3;										
QY	4	GGATCACTACATGCTGTGCGAGAGCGCCCTGGAGAGATTTAGAAATCTGCCAGTA	63							
Db	77	GGATCACTACATGCTGTGCGAGAGCGCCCTGGAGAGATTTAGAAATCTGCCAGTA	136							
QY	64	ATGGGGATGCACTAGAAACTTCCCAATGTCTACACACAGGCCCGCCCAACCCAG	123							
Db	137	ATGGGGATGCACTAGAAACTTCCCAATGTCTACACACAGGCCCGCCCAACCCAG	196							
QY	124	CCAGCCAAAGGAGGACGACACCAACCCCGCCCGCCAGAGACCTCCAAACCC	183							
Db	197	CCAGCCAAAGGAGGACGACACCAACCCCGCCCGCCAGAGACCTCCAAACCC	256							
QY	184	AAGAGGAGACCAACCACTGCAATACGTCCTCAGAGTGGTCTCAAGACACTAGGAA	243							
Db	257	AAGAGGAGACCAACCACTGCAATACGTCCTCAGAGTGGTCTCAAGACACTAGGAA	316							
QY	244	CACCACTTTGATGCGCTTTCCAGACGCTGTGATGCGCTCAAGCTGAACCTCCCTGAT	303							
Db	317	CACCACTTTGATGCGCTTTCCAGACGCTGTGATGCGCTGAACCTGACCTCCCTGAT	376							
QY	304	TACTATAGATTCATTAAGAGCCATATGATATGGGAACAATAAAGCCCTTGGAAAC	363							
Db	377	TACTATAGATTCATTAAGAGCCATATGATATGGGAACAATAAAGCCCTTGGAAAC	436							
QY	364	AACATTTACTGGAATGCTCAGGAATGTATCCAGAGCTTCAACACTATGTTTACAATGT	423							
Db	437	AACATTTACTGGAATGCTCAGGAATGTATCCAGAGCTTCAACACTATGTTTACAATGT	496							
QY	424	TACATCTACACAAGCCTGAGATGACATAGTCTTAATGCGAGAAGCTCTGGAAAAGCTC	483							
Db	497	TACATCTACACAAGCCTGAGATGACATAGTCTTAATGCGAGAAGCTCTGGAAAAGCTC	556							
QY	484	TTCTTGGAAAAATTAATGAGCTTACGACAGAAAGAACCGAGATCATGATGTCCAGGCA	543							
Db	557	TTCTTGGAAAAATTAATGAGCTTACGACAGAAAGAACCGAGATCATGATGTCCAGGCA	616							

QY 544 AAAGGAAAGAGACGCGGGAGGAAAAGAAACAGGGACAGCAAACTCGGCTTCCACGGTA 603  
 Db 617 AAAGGAAAGAGACGCGGGAGGAAAATATACAGTACAGCAAACTCGGCTTCCACGGTA 676  
 QY 604 CCAAACACACTCAAGATCGACTCTCTCCGCGAGACCCAGACCTCTAGCCGAATCTCTCT 663  
 Db 677 CCAAACACAACTCAAGACTCGACTCTCTCCGCGAGACCCAGACCTCTAGCCGAATCTCTCT 736  
 QY 664 CCTGTGCAGGCGCAGCCACCCCTTCCCTGCGCGTACACCCCGGACTCATATGCTCCAGACC 723、  
 Db 737 CCTGTGCAGGCGCAGCCCTCTACCCCTTCTCGCGGTACACCCCGGACTCATATGCTCCAGACC 796  
 QY 724 CCTGTCAATG-ACAGTGTGTGCTCTCCCGAGCCAC--TTCAGACGCGCCCGCGCAATGCGCCCGC 780  
 Db 797 CCTGTCAATGACAGTGTGTGCTCTCCCGAGCCACCTTGGAAACAGACCCCGGCAATGTGCGCCCGC 856  
 QY 781 CAGCGA--CAACCCCGCACCCGCTCGAGGCTCCCAAGCCCGGTACAGAGCACCACCACTATC 837  
 Db 857 CCACCCAGAGAAACCCCGCCCGGTTTGAAGATCTCCCAAGCGCCCGGACTGTAGAACCCCGCCACCC 915

RESULT	13
B0650970	
LOCUS	B0650970
DEFINITION	B0650970 926 bp mRNA linear EST 15-JUL-2002
	AGENCOURT 8488559 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:629607B
	5' , mRNA sequence.

VERSION BQ650970.1 GI:21775142

ORGANISM Homo sapiens

REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 926)
TITLE	NIH-MGC <a href="http://mgc.ncbi.nlm.gov/">http://mgc.ncbi.nlm.gov/</a> .
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT	Unpublished (1999)
	Contact: Robert Strausberg, Ph.D.

Tissue Procurement: CGAP (Stanford)  
 cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MCC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNLN at:  
<http://image.lnln.gov>  
 Plate: L10CM2503 row: f column: 15  
 High quality sequence stop: 595.

FEATURES	Location/Qualifiers
source	1. .926

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:6296078"
/clone_lib="NIH-MGC-100"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pOTB1; site.1: XhoI; Site.2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCGAGGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH-MGC
Library."

```

BASE COUNT	267 a	290 c	218 g	151 t
ORIGIN				

Query Match	31.4%:	Score	730.4:	DB	14:	Length	926:
Best Local Similarity	93.7%:	Pred. No.	3.1e-121:				
Matches	772:	Conservative	0:	Mismatches	51:	Indels	1:
						Gaps	1:

4 GGGATCACTAGCATGTCTGCGGAGAGCGGCCCTGGGACGAGATTGAGAAATCTGCCAGTA 63

Db	77	GGGATCTAGCATCTCTGCGGAGAGCGCCCTTGGGACGAGATTGAGAAATCTGCCAGTA	136
OY	64	ATGGGGGATGGACTGTGAAACCTTCCCAATGTCTACACACAGGCCAGGCCCAACCCAG	123
Db	137	ATGGGGGATGGACTGTGAAACCTTCCCAATGTCTACACACAGGCCAGGCCCAACCCAG	196
OY	124	CCAGCACAACGACGACGACACCAACCCCGCCCGCCAGAGACCTCCAAACCTTACAGGCC	183
Db	197	CCAGCACAACGACGACGACACCAACCCCGCCCGCCAGAGACCTCCAAACCAAGGCC	256
OY	184	AAGAGCGAGACCAACCAACTGCAATACCTGCTCAGAGTGGTCTCAAGACACTATGAAA	243
Db	257	AAGAGCGAGACCAACCAACTGCAATACCTGCTCAGAGTGGTCTCAAGACACTATGAAA	316
OY	244	CACGAGTTTGATGCGCTTTCGAGCAGCGCTGTGATGCGCTCAAGCTGAACCTCCCTGAT	303
Db	317	CACGAGTTTGATGCGCTTTCGAGCAGCGCTGTGATGCGCTCAAGCTGAACCTCCCTGAT	376
OY	304	TACTTAAAGATCATTTAAACGCTTATGATATATGGACAAATPAAAGACCGCTTGGAAAC	363
Db	377	TACTTAAAGATCATTTAAACGCTTATGATATATGGACAAATPAAAGACCGCTTGGAAAC	436
OY	364	AACCTTCTGSAATGCMTCAGSAATGATATCCAGAGCTTCAACACATATGTTTACAAATGT	423
Db	437	AACCTTCTGSAATGCMTCAGSAATGATATCCAGAGCTTCAACACATATGTTTACAAATGT	496
OY	424	TACATCTCAACACAGCCCTGAGAGTACATATGCTTAATGGAGAACTCTGGAAAAAGTC	483
Db	497	TACATCTCAACACAGCCCTGAGAGTACATATGCTTAATGGAGAAAGCTCTGGAAAAAGTC	556
OY	484	TTCTTTCGCAAAAAATTAATAGCTTACCCACAGAGAAGAACCGAGATCATATATGTCAGGCA	543
Db	557	TTCTTTCGCAAAAAATTAATAGCTTACCCACAGAGAAGAACCGAGATCATATATGTCAGGCA	616
OY	544	AAAGGAAGAGGACGCGGAGAGAAAGAAACAGAGAACCGAGGCTTTCACGGTA	603
Db	617	AAAGGAAGAGGACGCGGAGAGAAAGAAACAGAGAACCGAGGCTTTCACGGTA	676
OY	604	CCAAACACAGACTCAAGCATCTCTCCGAGACCCAGACCCCTCAGCCCAATCTCTCT	663
Db	677	CCAAACACAGACTCAAGCATCTCTCCGAGACCCAGACCCCTCAGCCCAATCTCTCT	736
OY	664	CCCTGTGACAGGCCGCTACGCCCTTCCCTCCGCTACGCCCGAGACCTTATGTCAGACC	723
Db	737	CCCTGTGACAGGCCGCTACGCCCTTCCCTCCGCTACGCCCGAGACCTTATGTCAGACC	796
OY	724	CCCTGTCAAT-GACAGTGGTGCCTCCCGACCCACATGACAGAGCCCGCGCAAGTGCCTCCCA	782
Db	797	CCCTGTCAATGACAGTGGTGCCTCCCGACCCCAATGAAAGAGCCCGCGCAAGAGGCC	856
OY	783	GCCACATACCCCAACCGCTCCAGCTCCCGACCGGTATACAGGCC	826
Db	857	CCGAGGTAAAAAACCCTCCACACCCGTTTAAGTTTCCCGCAGGCC	900

RESULT 14	BM983041/c	740 bp	mRNA	linear	EST 21-MAR-2002
LOCUS	BM983041				
DEFINITION	UI-CF-EN1-act-f-11-0-UI.s1 UI-CF-EN1 Homo sapiens cDNA clone UI-CF-EN1-act-f-11-0-UI 3', mRNA sequence.				

ACCESSION	BM983041
VERSION	BM983041.1
	GI:19607155

<b>KEYWORDS</b>	<b>EST.</b>
<b>SOURCE</b>	human.

ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 740)  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo

AUTHORS	Bonaldo, M.F., Lennon, G. and Soares, M.B.
TITLE	Normalization and subtraction: two approaches to facilitate gene

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477  
 COMMENT Contact: McCray, PB  
 McCray Lab  
 University of Iowa  
 2024 University of Iowa Med Labs, Iowa City, IA 52242, USA  
 Tel: 319 356 4866  
 Fax: 319 356 7171  
 Email: paul-mccray@uiowa.edu  
 Tissue Procurement: Dr. M. J. Welsh, University of Iowa  
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).  
 Seq primer: M13 FORWARD  
 POLYA-Yes

FEATURES  
 source location/Qualifiers  
 1. 740  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="UI-CF-EN1-act-f-11-0-UI"  
 /clone\_lib="UI-CF-EN1"  
 /tissue\_type="Primary Lung Cystic Fibrosis Epithelial Cells"  
 /dev\_stage="Adult"  
 /lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
 /note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site.1: EcoR I; Site.2: Not I; UI-CF-EN1 is a normalized cDNA library containing the following tissue(s): Primary Lung Cystic Fibrosis Epithelial Cells. The library was constructed according to Bonaldi, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CTCCTCAGGT.  
 TAG\_LIB=UI-CF-EN1  
 TAG\_TISSUE=Human Lung Epithelial Cell lines untreated LPS 6hr to LPS 24h  
 TAG\_SEQ=CTCTCAGGT"  
 BASE COUNT 115 a 198 c 240 g 184 t 3 others  
 ORIGIN  
 Query Match 31.3%; Score 729; DB 14; Length 740;  
 Best Local Similarity 98.9%; Pval. No. 5.0e-121;  
 Matches 732; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 901 ACCCCACACACATGACACCATTCAGACAGACCTCGCCCGGAGCCCAAGAC 960  
 |||||||  
 Db 740 ACCCCACACACATGACACCATTCAGACAGACCTCGCCCGGAGCCCAAGAC 681  
 |||||||  
 QY 961 ACCAAGCTGGCGCAGCGCGGAGAGCAGCGCGCTTGAAACCTCCAAAGAGAGCTG 1020  
 |||||||  
 Db 680 ACCAAGCTGGCGCAGCGCGGAGAGCAGCGCGCTTGAAACCTCCAAAGAGAGCTG 621  
 |||||||  
 QY 1021 CCCGACTCTAGACAGACCCAGACAGAGAGAGCAAGGCTCTGGAGCAGCTCAAG 1080  
 |||||||  
 Db 620 CCCGACTCTAGACAGACCCAGACAGAGAGAGCAAGGCTCTGGAGCAGCTCAAG 561  
 |||||||  
 QY 1081 TGCTGAGCGGCATCTCAAGAGAGATGTTTGCCAAAGACCGCGCTACGCTGGGCC 1140  
 |||||||  
 Db 560 TGCTGAGCGGCATCTCAAGAGAGATGTTTGCCAAAGACCGCGCTACGCTGGGCC 501  
 |||||||  
 QY 1141 TTCTACAAGCGCTGTGAGAGCTGGGCGCTACACGACTCTGTGACATCATCAAG 1200  
 |||||||  
 Db 500 TTCTACAAGCGCTGTGAGAGCTGGGCGCTACACGACTCTGTGACATCATCAAG 441  
 |||||||  
 QY 1201 CACCCCATGAGACACACATCAAGTCTAACTGGAGGCCCTGTGATCCGTGATGCT 1260  
 |||||||

Db 440 CACCCCATGAGACACACATCAAGTCTAACTGGAGGCCCTGTGATCCGTGATGCT 381  
 |||||||  
 QY 1261 CAGAGATTGGTGTGAGAGCTCCGATTTGTTCTCAACTGCTTAAGTACAACTCTCT 1320  
 |||||||  
 Db 380 CAGAGATTGGTGTGAGAGCTCCGATTTGTTCTCAACTGCTTAAGTACAACTCTCT 321  
 |||||||  
 QY 1321 GACCATGAGGTGTGAGCCATGAGCCCGCAAGCTCCAGATGTGTTTGAATGCTTTGCC 1380  
 |||||||  
 Db 320 GACCATGAGGTGTGAGCCATGAGCCCGCAAGCTCCAGATGTGTTTGAATGCTTTGCC 261  
 |||||||  
 QY 1381 AAGATGCCGAGACAGCTGTGAGAGACCAAGTGTGCTCTCCCGGACAGTCCCTCT 1440  
 |||||||  
 Db 260 AAGATGCCGAGACAGCTGTGAGAGACCAAGTGTGCTCTCCCGGACAGTCCCTCT 201  
 |||||||  
 QY 1441 CCCACCAAGTTGTGGCCCGCCCTCATCCAGCAGACAGCAGATGATGCTCCGGAC 1500  
 |||||||  
 Db 200 CCCACCAAGTTGTGGCCCGCCCTCATCCAGCAGACAGCAGATGATGCTCCGGAC 141  
 |||||||  
 QY 1501 AGTACAGTTCTGACTGATGACTGTGAGAGAGAGCGAGCCAGCGGTGAGCTTCAG 1560  
 |||||||  
 Db 140 AGTACAGTTCTGACTGATGACTGTGAGAGAGAGCGAGCCAGCGGTGAGCTTCAG 81  
 |||||||  
 QY 1561 GAGCAGCTCAAAAGCCGTGACAGACAGCTTGCACCTCTCTAGCCCCAGACAA 1620  
 |||||||  
 Db 80 GAGCAGCTCAAAAGCCGTGACAGACAGCTTGCACCTCTCTAGCCCCAGACAA 21  
 |||||||  
 QY 1621 CCAAGAAAAAGAGAAAGA 1640  
 |||||  
 Db 20 CCAAGAAAAAGAGAAAGA 1

RESULT 15  
 B0654353  
 LOCUS  
 DEFINITION B0654353 938 bp mRNA linear EST 15-JUN-2002  
 ACCESSION AGENOUR1\_8286504 NIH\_MGC\_100 Homo sapiens cDNA clone IMAGE:629391  
 B0654353  
 VERSION 5, mRNA sequence.  
 KEYWORDS B0654353.1 GI:21778525  
 EST  
 SOURCE  
 ORGANISM  
 human.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 938)  
 NIH-MGC http://mgc.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 unpublished (1999)  
 CONTACT Robert Strausberg, Ph.D.  
 Email: cga@bbs.fema11.nih.gov  
 Tissue Procurement: CGAP (Stanford)  
 cDNA Library Preparation: Rubin Laboratory  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.lnl.gov  
 Plate: LNCM2512 row: h column: 24  
 High quality sequence stop: 680.

FEATURES  
 source location/Qualifiers  
 1. 938  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:629391"  
 /clone\_lib="NIH\_MGC\_100"  
 /tissue\_type="hepatocellular carcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: Liver; Vector: pOT73; Site.1: XhoI; Site.2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using Zap-cDNA synthesis kit (Stratagene) and Superscript

II RT (Life Technologies). Note: this is a NTR\_MGC  
Library.  
BASE COUNT 262 a 301 c 225 g 150 t  
ORIGIN

Query Match 30.7%; Score 715.2; DB 14; Length 938;  
Best Local Similarity 91.6%; Pred. No. 1.7e-118;  
Matches 779; Conservative 0; Mismatches 68; Indels 3; Gaps 2;

OY 4 GGGATCATAGATGTCCTCGGAGAGCGGCGCTGGGAGCAGATTGAGAAATCTGCCAGTA 63  
DB 77 GGGATCATAGATGTCCTCGGAGAGCGGCGCTGGGAGCAGATTGAGAAATCTGCCAGTA 136  
OY 64 ATGGGGGATGAGCTAGAAAATTCTCCAAATGTCTACAAACAGAGCCCGAGGCCCAACCCAG 123  
DB 137 ATGGGGGATGAGCTAGAAAATTCTCCAAATGTCTACAAACAGAGCCCGAGGCCCAACCCAG 196  
OY 124 CGAGCCCAAGCGCCAGCAACCCCGCCCGCCGAGACCTCCAAACCTTAACAAAGCCC 183  
DB 197 CGAGCCCAAGCGCCAGCAACCCCGCCCGCCGAGACCTCCAAACCTTAACAAAGCCC 256  
OY 184 AAGAGGAGACCAACCACTGCATCTGCTCAGAGTGTGCTCAAGACACTATGAGAA 243  
DB 257 AAGAGGAGACCAACCACTGCATCTGCTCAGAGTGTGCTCAAGACACTATGAGAA 316  
OY 244 CACCACTTTCATGGCTTTCAGACAGCTGTGATGCCGTCAAGCTGAACCTCCCTGAT 303  
DB 317 CACCACTTTCATGGCTTTCAGACAGCTGTGATGCCGTCAAGCTGAACCTCCCTGAT 376  
OY 304 TACTATTAAGATCAATTAAGAGCTATGAGTATGGGAAACATTAAGAGGCTTGGAAAC 363  
DB 377 TACTATTAAGATCAATTAAGAGCTATGAGTATGGGAAACATTAAGAGGCTTGGAAAC 436  
OY 364 AACTATTACTGAGATGCTCAGATATGATCAGAGCTTCAACACTATGTTTACAATTTG 423  
DB 437 AACTATTACTGAGATGCTCAGATATGATCAGAGCTTCAACACTATGTTTACAATTTG 496  
OY 424 TACATCTACACAAAGCTTGAGATGACATAGCTTATATGGCAGAGCTCTGGAAGATTC 483  
DB 497 TACATCTACACAAAGCTTGAGATGACATAGCTTATATGGCAGAGCTCTGGAAGATTC 556  
OY 484 TTTCTGCAAAAATTAATAGCTACCCACAGAAACCGAGATCATATGATGTCAGGCA 543  
DB 557 TTTCTGCAAAAATTAATAGCTACCCACAGAAACCGAGATCATATGATGTCAGGCA 616  
OY 544 AAAGGAAGAGAGCTGGGAGGAAAGAAACAGGGACAGCAAAAACCTGGCTTCCACGTA 603  
DB 617 AAAGGAAGAGAGCTGGGAGGAAAGAAACAGGGACAGCAAAAACCTGGCTTCCACGTA 676  
OY 604 CCAAAACACACTCAGCATGACTCTCCGACAGACCCAGACCCCTCAGCCGAAATCTCT 663  
DB 677 CCAAAACACACTCAGCATGACTCTCCGACAGACCCAGACCCCTCAGCCGAAATCTCT 736  
OY 664 CCTGTGCAAGGCGCTACCCCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 722  
DB 737 CCTGTGCAAGGCGCTACCCCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 796  
OY 723 CCTGTGATGA--CAGTGTGCTCCCGACGCACTCAGAGCGCCCGCCAGATGCCCCC 780  
DB 797 CCTGTGATGAAGAGTGGGTGCTCCCGACGCACTGAGAGCGCCCGCCAGATGCCCCC 856  
OY 781 CAGCCCAAAACCCCGCTCAGCTCCCGACGCCGCTACAGAGCCACCCACCCATCATTC 840  
DB 857 CCCCCAGGCCCAAAACCCCGCTCAGCTCCCGACGCCGCTACAGAGCCCGGTAAGAGCCC 916  
OY 841 GCGGCCAACC 850  
DB 917 CCCCCCCCCC 926

Search completed: March 1, 2003, 14:40:45  
Job time : 3262 secs